

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:38:36 ; Search time 7810.03 Seconds
(without alignments)
-7.776 Million cell updates/sec

Title: US-08-956-991-6

Perfect score: 20
Sequence: 1 cctgattgacctgcaggaag 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_dal.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vl.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vl.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	20	100.0	6110	11 AF023449	AF023449 Homo sapi
2	20	100.0	6413	11 AF023450	AF023450 Homo sapi
3	20	100.0	145861	11 AF064862	AF064862 Homo sapi
4	20	100.0	159424	11 AF064865	AF064865 Homo sapi
5	17.4	87.0	53718	43 AC013576	AC013576 Homo sapi
6	17.4	87.0	89948	11 HS756623	AL035661 Human DNA
7	17.4	87.0	106319	11 HS74714	AL009051 Homo sapi
8	17.4	87.0	116967	41 AC009389	AC009389 Drosophila
9	17	85.0	139134	45 AC011140	AC011140 Homo sapi
10	16.8	84.0	3374	9 HSMYOH	Y00821 Human RNA
11	16.8	84.0	4060	12 MUSAE3A	M28383 Mouse AEs3 m
12	16.8	84.0	4350	12 RN035022	U35022 Rattus norv
13	16.8	84.0	4437	9 HSBMHC	X51592 Human MHC m
14	16.8	84.0	5259	9 HUMDMP	J05211 Human Desmo
15	16.8	84.0	5866	3 AB025262	AB025262 Sus scrof
16	16.8	84.0	5925	40 AF111785	AF111785 Homo sapi
17	16.8	84.0	5929	3 AB025261	AB025261 Sus scrof
18	16.8	84.0	5932	3 AB025260	AB025260 Sus scrof
19	16.8	84.0	5956	11 AF111784	AF111784 Homo sapi
20	16.8	84.0	6010	5 HSMYOSIN	Z38133 H.sapiens m
21	16.8	84.0	6016	11 HUMMDPI	M36769 Homo sapien
22	16.8	84.0	9588	40 HUMMDPI	M77830 Homo sapien
23	16.8	84.0	33593	6 AF063097	AF063097 Bacteriop
24	16.8	84.0	64356	11 HS512B11	AL031058 Human DNA
25	16.8	84.0	104913	11 AC004242	AC004242 Homo sapi
26	16.8	84.0	125669	41 AC011285	AC011285 Homo sapi
27	16.8	84.0	125937	11 HS69M21	AL031735 Human DNA
28	16.8	84.0	198582	11 AC005291	AC005291 Homo sapi
29	16.8	84.0	204082	11 AC001342	AC005333 Homo sapi
30	16.8	84.0	225747	41 AC008083	AC008083 Homo sapi
31	16.4	82.0	758	11 HSLRPS525	AF058421 Homo sapi
32	16.4	82.0	1284	11 AF022150	AF022150 Homo sapi
33	16.4	82.0	1284	11 HSAF000972	AF000972 Homo sapi
34	16.4	82.0	1982	11 AF033021	AF033021 Homo sapi
35	16.4	82.0	2238	11 AF022797	AF022797 Homo sapi
36	16.4	82.0	28690	41 AC009072	AC009072 Homo sapi
37	16.4	82.0	30543	41 AC008618	AC008618 Homo sapi
38	16.4	82.0	42709	41 AC011520	AC011520 Homo sapi
39	16.4	82.0	103194	8 ATAC006223	AC006223 Arabidops
40	16.4	82.0	126956	11 HS8B1	AL053701 Human DNA
41	16.4	82.0	166465	32 HS1447E21	AL050336 Homo sapi
42	16	80.0	2875	12 MUSMHOAB	M20985 Mouse MHC c
43	16	80.0	2945	12 AF016308	AF016308 Mus muscu
44	16	80.0	45000	43 AC013257	AC013257 Leishmani
45	16	80.0	65386	43 AC015461	AC015461 Homo sapi

ALIGNMENTS

RESULT 1
AF023449 AF023449 6110 bp mRNA PRI 01-JUN-1998
LOCUS
DEFINITION Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)
ACCESSION AF023449
VERSION AF023449.1 GI:3169765

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 6110)
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,
Lyons, G.E. and Korenberg, J.R.
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a
Down Syndrome Region and is Involved in the Development of the
Nervous System
Unpublished
2 (bases 1 to 6110)
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,
Lyons, G.E. and Korenberg, J.R.
Direct Submission
Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los
Angeles, CA 90048-1869, USA
Location/Qualifiers
1. 6110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21 (trisomy 21)"
/map="21q22, between HMGI4 and MXI"
/tissue_type="brain"
/dev_stage="14 weeks, fetal"
/clone="CHD2-42"
/note="derived from alternately-spliced mRNA"
1. 6110
/gene="DSCAM"
/note="member of immunoglobulin superfamily; involved in
nervous system development"
/product="Down syndrome cell adhesion molecule"
/protein_id="AAC17966.1"
/db_xref="GI:3169766"
/translacion="VESEDLSLSLYEVNASLQEVFASTGTGLVPCPAAGIPVTLRM
/LATGEITDVPGRVHHPNGTLOIPPEPPSSSTLIHNTYCTAENPSGKRSQDV
HITAVIREPVTVEREDOKTRMGVAVAFKCIIPSSVEAFITVVMERDQVSIYSGSFL
ITISGALYIKDVONEDGLYRCITRRTTGETRQNSALPEVSDANAPSLIDFED
HRKMAQORVELPCALGHPEDPRWKDMPLLSGRQKVTGTLINIRSDSGS
YCEVSNRYGTAVYIGRLVYKOPKATISPRKVSXSGSOVSLSCSTGTEDDELAMY
RNGELINPGKNRITGIGNHENLIMDHVYKSDGAYOCFYKDKLSADVOYVLEDEGT
PKIIISAFSEKVVSPAEVSLMCNVKGPPLTITWTLDDELKSGSHRISMTSGN
VSLYINSSOVBDGQVYRGTANNSAGVLYOARINRGPAISIRPKMTITAIAGRTY
IHCRTVGYPIYSIKYKNSKLLPFRNOYAFENNGTLKLSVOKEVDECEITCNVVO
POLSTQSVHVTAKVPPTIOPEEPERSIGQVFIQVAVSGDLPTITWQDGRIP
GSLGVTIDNIDFTSLRIKSLMHNGNTICIAENAAVEHQSQLIVVPRFVVO
RDDGIGYKAVILNCSAEGYPTIYKFSKAGQPOPIALNGRIQVLSGLIK
HVEEDGYGLCKVNSGADVSKSMILYKIPAMITSPTNTLATQSKREKSCAH
GEMPIYRWKEDRIINPEMARLYSTKEGVEVSTIOLIPVREDSPFCFKANS
YGDGRILOITQVPEPPEIETDKVAKRITITRMWGDGNSPTITGIETCKNSDS
WDSQRTKDVSPQNSKTTIDIPSSITSIRMTAKNRIGKSEPSNELITIAEAAVDG
PPOEVHLEPISSOIRVTKAPRKLQNGIIRGOYIGREYSTGNGFONIIISVDSG
DSEVITLDNLNKTFOGLVVOACNRAGTGPSSOEITITLEDVSPSPENVAITSP
ESISISWSTLSKALNGILOGFVVIYMANLMDEIGEIKNITTOPLSELDEKXTN
YSIOVLAFTRAGDGVSEIOIFRTKEDVPGPAGVAKAAASAMVSVMLPPLKNGI
IRKTYVCSHPYPIVISEPASPSFSYRPNLSRNOYSYWVAVNTSGRGNSSII
TVERPLAKAPRILITFSTYTPPMKQIVLPCKARQSPFVAVKMKOSNPTLITDIDG
RRIISFNSGSIITFVKAEDSGYSCIANNNWSDGELILNQVVPDQRLVLSKITS
SSITLWLPDNGSSIRGYILIOYSEDSNOMGSPISPERSYRLENLNCGTWKT
SLTQNGVGPRISEIIEAKTLKEPOFSKOEOLFASINTRVNLIGNNDGCGPITS
NYTLEVPFGTWTAAORTSLSKSYLLYDQEAETVLOMRVNSAGCAKKAANRQAL
FNGSGTIPLIKSVONEGLTNEGKMLVITSCILVGLVLLFVLLIYVRRRRROR
LKLRLAKSLAEMLAKNRTSTLTLSKOOTLBMHIDIPRAOLILEERTWTITDIDS
TVLLIDADEEAKOKSLVTHVYOSVQATPLVDVSDAPGNTPTRRNAGA
TANRRAQWTLNRPHTISAHTLITTDWRLLPRAAGSVKESDYSVSPQDTRAR
SNVYSASTYEELARAEVAKMEOLRHAFTITECFISDTSSEQLTACTNETYD
SLTSTPSESGICRFTASPKPDGGRVMNMAVPAKAIQGVTSYICLHTELMTEFC"

BASE COUNT 1621 a 1658 c 1548 g 1283 t

ORIGIN

Query Match 100.0%; Score 20; DB 11; Length 6110;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgtgtgaccctgcaggaag 20
|||||
Db 4560 CCTGTATGACTGACGAGAG 4579

RESULT 2
AF023450
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 6413)
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,
Lyons, G.E. and Korenberg, J.R.
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a
Down Syndrome Region and is Involved in the Development of the
Nervous System
Unpublished
2 (bases 1 to 6413)
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,
Lyons, G.E. and Korenberg, J.R.
Direct Submission
Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los
Angeles, CA 90048-1869, USA
Location/Qualifiers
1. 6413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21 (trisomy 21)"
/map="21q22, between HMGI4 and MXI"
/tissue_type="brain"
/dev_stage="14 weeks, fetal"
/clone="CHD2-52"
/note="derived from alternately-spliced mRNA"
1. 6413
/gene="DSCAM"
/note="member of immunoglobulin superfamily; involved in
nervous system development"
/product="Down syndrome cell adhesion molecule"
/protein_id="AAC17967.1"
/db_xref="GI:3169768"
/translacion="MMILASLQSFANVSEDLHSLYEVNASLQEVFASTGTGLV
PCPAAGIPVTLRM/LATGEITDVPGRVHHPNGTLOIPPEPPSSSTLIHNTYCTA
ENPSGKRSQDVHITAVIREPVTVEREDOKTRMGVAVAFKCIIPSSVEAFITVVM
ERDQVSIYSGSFLITISGALYIKDVONEDGLYRCITRRTTGETRQNSALPEVSDAN
APSLIDFEDHRKMAQORVELPCALGHPEDPRWKDMPLLSGRQKVTGTLINIRSDSGS
YCEVSNRYGTAVYIGRLVYKOPKATISPRKVSXSGSOVSLSCSTGTEDDELAMY
RNGELINPGKNRITGIGNHENLIMDHVYKSDGAYOCFYKDKLSADVOYVLEDEGT
PKIIISAFSEKVVSPAEVSLMCNVKGPPLTITWTLDDELKSGSHRISMTSGN
VSLYINSSOVBDGQVYRGTANNSAGVLYOARINRGPAISIRPKMTITAIAGRTY
IHCRTVGYPIYSIKYKNSKLLPFRNOYAFENNGTLKLSVOKEVDECEITCNVVO
POLSTQSVHVTAKVPPTIOPEEPERSIGQVFIQVAVSGDLPTITWQDGRIP
GSLGVTIDNIDFTSLRIKSLMHNGNTICIAENAAVEHQSQLIVVPRFVVO
RDDGIGYKAVILNCSAEGYPTIYKFSKAGQPOPIALNGRIQVLSGLIK
HVEEDGYGLCKVNSGADVSKSMILYKIPAMITSPTNTLATQSKREKSCAH
GEMPIYRWKEDRIINPEMARLYSTKEGVEVSTIOLIPVREDSPFCFKANS
YGDGRILOITQVPEPPEIETDKVAKRITITRMWGDGNSPTITGIETCKNSDS
WDSQRTKDVSPQNSKTTIDIPSSITSIRMTAKNRIGKSEPSNELITIAEAAVDG
PPOEVHLEPISSOIRVTKAPRKLQNGIIRGOYIGREYSTGNGFONIIISVDSG
DSEVITLDNLNKTFOGLVVOACNRAGTGPSSOEITITLEDVSPSPENVAITSP
ESISISWSTLSKALNGILOGFVVIYMANLMDEIGEIKNITTOPLSELDEKXTN
YSIOVLAFTRAGDGVSEIOIFRTKEDVPGPAGVAKAAASAMVSVMLPPLKNGI
IRKTYVCSHPYPIVISEPASPSFSYRPNLSRNOYSYWVAVNTSGRGNSSII
TVERPLAKAPRILITFSTYTPPMKQIVLPCKARQSPFVAVKMKOSNPTLITDIDG
RRIISFNSGSIITFVKAEDSGYSCIANNNWSDGELILNQVVPDQRLVLSKITS
SSITLWLPDNGSSIRGYILIOYSEDSNOMGSPISPERSYRLENLNCGTWKT
SLTQNGVGPRISEIIEAKTLKEPOFSKOEOLFASINTRVNLIGNNDGCGPITS
NYTLEVPFGTWTAAORTSLSKSYLLYDQEAETVLOMRVNSAGCAKKAANRQAL
FNGSGTIPLIKSVONEGLTNEGKMLVITSCILVGLVLLFVLLIYVRRRRROR
LKLRLAKSLAEMLAKNRTSTLTLSKOOTLBMHIDIPRAOLILEERTWTITDIDS
TVLLIDADEEAKOKSLVTHVYOSVQATPLVDVSDAPGNTPTRRNAGA
TANRRAQWTLNRPHTISAHTLITTDWRLLPRAAGSVKESDYSVSPQDTRAR
SNVYSASTYEELARAEVAKMEOLRHAFTITECFISDTSSEQLTACTNETYD
SLTSTPSESGICRFTASPKPDGGRVMNMAVPAKAIQGVTSYICLHTELMTEFC
ATQOKKEMSGTAKGKPIITVIRWEKEDRIINPEMARLYSTKEGVEVISTIOILPTV

BASE COUNT
ORIGIN

1633 a 1779 c 1709 g 1292 t

REDSCFFSCAHINSGEDRCIGTOLTVOPEPPPELEIKDKVAKARTITLMTMGEPGNBR
ITGIDIECKNRKSDMSAQRTKDVSPQJNLSKTIIDHPSSTIIRGAKINAKNGISEPS
NELTITDEAPADBPPOEVLHLEPISOSI RTWMAKHLONGIIRGVOISREYSTVG
GNFENITSVDTSGSEVYITLDNLNKFTQYLVAQACNRKATGCSOEIITITLEDVP
SYEPENVOAIVTSPESISISMTLSKLEKLNILLOGFRIYWANIMDGLGEIKINITTT
OPSLDGLKENTYKSTOYLATFACGAGVRESOIFETTKEDVPPPGVAKAAASAM
VFSWMLPPLKLNGIIRKTYVCSHPPIVISEFMA SPDSSEYI RPNLSRNOISYVWY
AVTSAGRGNSSEIITVEPLANAPARILTFSTVITVPMKDIATLCKAAAGDSRPAVKM
KDSNGTPLSVLTIDGRSRIFSNGSFIIRVAKNDEGGYSICANNMGWSEIILNLQVOY
PPDOPRLTIVSKTISSTITLWLPDNGNSIRGYILOESDNOMSGSFIIPSPERSR
RLEUKCGCTWKTFTLQNGVGPRIISIIIAKTLKGPOSKDELFASINTTRVRL
NLIMDNGGCPITSFETLEYPFGTTVMTTARTLSNSYILYKDQEWATWYELQARVNCN
SAGCAEROKAEANRCKERS"

Query Match	100.0%	Score 20	DB 11	Length 6413
Best Local Similarity	100.0%	Pred. No. 1.7		
Matches 20	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy	1	ccgtgatgacctgcaggaag	20
Db	5054	ccgtgatgacctgcaggaag	5073

RESULT	3		
AF064862			
LOCUS	AF064862	145861 bp	DNA
DEFINITION	Homo sapiens chromosome 21q22.3 PAC 31p10, complete sequence.		

VERSION AF064862.1 GI:3171157
KEYWORDS HTG.
SOURCE human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145861)

AUTHORS
Taudien, S., Nordsiek, G., Hildmann, T., Dagand, E., Drescher, B.,
Weber, J., Schattevoy, R., Yaspo, M.-L. and Rosenthal, A.

JOURNAL Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES Location/Qualifiers
SOURCE 1. 145861

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
```

```

complement (224. .380)
/note="GRAIL, score = 59.000%, comment = good"
/evidence=not experimental

```

```
repeat_region 240: 1474
                /rpt_family="MIR"
                /evidence=not_experimental
                complement(1011..1160)
repeat_region
```

```
repeat_region
1175. 1295
/rpt_fam11v="L1MA9"
/ape_data=not_experimental
```

```

exon
1808. .1963
/feature=not_experimental
/note="Xpound exon predict
/evidence=not_experimental

```

```
each
1000
1997
1000
/note="GenScan, score = 1.86%, comment = Internal-exon 156
bp frame= 0 phase= 0"
evidence=not experimental
```

```

exon
1808. .1963
/note="M2EF, score = 67.98
/evidence=not experimental)

```

```

exon
1843.1963
/note="GRAIL, score = 100.000%, comment = excellent"

```

```

/evidence-not_experimental
complement(1908..2053)
/note="RALT, score = 53.000%, comment = good shadow"
/evidence-not_experimental

```

```

/rpt_family="MIR"
/evidence=not_experimental
3969. 4067
exon

```

```
bp frame: 1 phase: 0"
/evidence-not_experimental
3969. .4067
exon
```

```

/evidence=not_experimental
3969.4067
/note="MZF, score = 57.3%"

```

```

exon      3981.  .4075
/note="Xpound exon prediction, score = 82% (0%)"
/evidence=not_experimental

```

```

/rpt_family="Alusx"
/evidence=not_experimental
repeat_region 5103. .5382

```

```
/rpt_type=tandem
/rpt_unit=cgtagtcatagtcttcagtcgcatgacatatctcttgc
/evidence=not_experimental
```

```

/note="GRAIL, score = 48.000%, comment = marginal shadow"
/evidence=not_experimental
complement(5886..7061)
repeat_region

```

```
repeat_region /evidence=not_experiment  
complement(7186..7224)  
/rpt_family="MLTIE"
```

```
repeat_region complement(7467..7531)
              /ipt_family="MLTIE"
              /evidence=not_experimental
```

```

/rpt_family="L2"
/evidence=not_experimental
complement(8074. .8249)
repeat_region

```

```

evidence-not_experimental
complement(8465.8553)
/note="MZF1, score = 73.98"
exon

```

```
repeat_region complement(8606, .8854)
/rpt_family="MIR"
/evidence=not experimental
```

```

/note="M2EF, score = 95.78"
/evidence=not_experimental
complement(10335..10527)

```

```

/existence=not_experimental
10809..10921
/note="GRAIL, score = 52.000%, comment = good shadow"
exon

```

```
repeat_region
complement(10967..11374)
/rpt_family="MLTIB"
/evidence=not experimental
```

```

11488, 11621
/note="Xpound exon prediction, score = 628 (38)"
/evidence=not_experimental

```

```

/ncbi_molmap_score = 200
/evidence=not_experimental
11494..11621
/ncbi_gencode_score = 7588 comment = Internal exon 128
exon

```

exon /evidence-not_experimental
 11509. .11621
 /note="Xpound exon prediction, score = 74% (0%)"
 /evidence-not_experimental
 11535. .11621
 /note="GRAIL, score = 58.000%, comment = good shadow"
 /evidence-not_experimental
 12393. .12480
 /rpt_family="L2"
 repeat_region /evidence-not_experimental
 complement(12492. .12869)
 /rpt_family="L1MB3"
 repeat_region /evidence-not_experimental
 13831. .14128
 /rpt_family="AlusX"
 /evidence-not_experimental
 15597. .15646
 /rpt_family="L2"
 repeat_region /evidence-not_experimental
 complement(15672. .15819)
 /note="MZF, score = 53.3%"
 /evidence-not_experimental
 15717. .15762
 /rpt_family="PABLA"
 repeat_region /evidence-not_experimental
 complement(17207. .17470)
 /rpt_family="AlusX"
 /evidence-not_experimental
 complement(17433. .17534)
 /note="GRAIL, score = 87.000%, comment = excellent"
 /evidence-not_experimental
 18789. .19089
 /rpt_family="AluY"
 repeat_region /evidence-not_experimental
 complement(19490. .20236)
 /rpt_family="L1PA15"
 repeat_region /evidence-not_experimental
 complement(20244. .26377)
 /rpt_family="L1PA2"
 exon /evidence-not_experimental
 complement(20451. .22019)
 /note="GRAIL, score = 92.000%, comment = excellent"
 /evidence-not_experimental
 21819. .21970
 /note="MZF, score = 65.5%"
 exon /evidence-not_experimental
 complement(22035. .22267)
 /note="GRAIL, score = 59.000%, comment = good"
 /evidence-not_experimental
 22537. .22859
 /note="GRAIL, score = 49.000%, comment = marginal"
 exon /evidence-not_experimental
 complement(23004. .24272)
 /note="GRAIL, score = 87.000%, comment = excellent"
 /evidence-not_experimental
 26405. .27896
 /rpt_family="L1PA15"
 repeat_region /evidence-not_experimental
 complement(26551. .26723)
 /note="GRAIL, score = 50.000%, comment = good"
 /evidence-not_experimental
 26651. .26773
 /note="MZF, score = 52.1%"
 exon /evidence-not_experimental
 27887. .29472
 /rpt_family="L1P"
 repeat_region /evidence-not_experimental
 28426. .28580
 /note="MZF, score = 63.6%"
 exon /evidence-not_experimental
 29473. .29835
 /rpt_family="THE1B"
 repeat_region /evidence-not_experimental

repeat_region 29836. .29969
 /rpt_family="L1P"
 /evidence-not_experimental
 complement(30121. .30486)
 repeat_region /rpt_family="THE1C"
 /evidence-not_experimental
 complement(30940. .31299)
 /rpt_family="L1MC1"
 /evidence-not_experimental
 Query Match 100.0%; Score 20; DB 11; Length 145861;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ccggtatgacctgcaggaag 20
 Db 56986 CCTGTATGACCTCGAGAG 57005
 RESULT 4
 AF064865/c AF064865 159424 bp DNA PRI 02-JUN-1998
 LOCUS Homo sapiens chromosome 21q22.3 PAC 58D10, complete sequence.
 DEFINITION AF064865
 ACCESSION AF064865.1 GI:3171160
 VERSION HTG.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 159424)
 Blechschmidt,K., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
 Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
 Direct Submission
 TITLE Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
 Biology, Beutenbergstrasse 11, Jena 07745, Germany
 JOURNAL Location/Qualifiers
 FEATURES
 source
 1. .159424
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.3"
 repeat_region complement(1224. .1523)
 /rpt_family="AluY"
 /evidence-not_experimental
 1263. .1386
 /note="GC score = 8.60 (124bp)"
 /note="Region: GC content"
 exon /evidence-not_experimental
 1932. .2047
 /note="MZF, score = 92%"
 /evidence-not_experimental
 3299. .3348
 /note="Xpound exon prediction, score = 72% (0%)"
 /evidence-not_experimental
 complement(3308. .3614)
 /note="Genscan, score = -0.41%, comment = Terminal-exon
 107 bp frame: 0 phase: 2"
 /evidence-not_experimental
 3843. .4124
 /note="GRAIL, score = 53.000%, comment = good shadow"
 /evidence-not_experimental
 complement(3913. .4089)
 /note="Xpound exon prediction, score = 94% (0%)"
 exon /evidence-not_experimental
 complement(3913. .4066)
 /note="MZF, score = 72%"
 exon /evidence-not_experimental
 complement(3913. .4062)

exon /evidence=not_experimental
26464. .26634

Query Match 100.0%; Score 20; DB 11; Length 159424;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cctgtatgactgcaggaag 20
|||||
Db 7672 CCTGTATGACTGCAGGAG 7653

RESULT 5
AC013576/c
LOCUS
DEFINITION Homo sapiens chromosome 4 clone 364_L_4 map 4, LOW-PASS SEQUENCE
SAMPLING.
AC013576 53718 bp DNA HTG 13-NOV-1999
AC013576.1 GI:6403685
HTG: HTGS_PHASE0.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 (bases 1 to 53718)
JOURNAL Homo sapiens chromosome 4, clone 364_L_4
AUTHORS Unpublished
2 (bases 1 to 53718)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckhagalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearlilano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardys,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lenoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,R.,
McKwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.
* NOTE: This record contains 70 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved. 737: contig of 737 bp in length
* 1
* 738 1529: gap of unknown length
* 1530 2272: gap of unknown length
* 2273 3041: gap of unknown length
* 3042 3795: gap of unknown length
* 3796 4567: gap of unknown length
* 4568 5301: contig of 734 bp in length

* 5302 6044: gap of unknown length
* 6045 6809: gap of unknown length
* 6810 7559: gap of unknown length
* 7560 8331: gap of unknown length
* 8332 9068: gap of unknown length
* 9069 9894: gap of unknown length
* 9895 10600: gap of unknown length
* 10601 11360: gap of unknown length
* 11361 12088: gap of unknown length
* 12089 12904: gap of unknown length
* 12905 13665: gap of unknown length
* 13666 14424: gap of unknown length
* 14425 15198: gap of unknown length
* 15199 15951: gap of unknown length
* 15952 16705: gap of unknown length
* 16706 17459: gap of unknown length
* 17460 18215: gap of unknown length
* 18216 18996: gap of unknown length
* 18997 19777: gap of unknown length
* 19778 20514: gap of unknown length
* 20515 21378: gap of unknown length
* 21379 22154: gap of unknown length
* 22155 22915: gap of unknown length
* 22916 23653: gap of unknown length
* 23654 24412: gap of unknown length
* 24413 25203: gap of unknown length
* 25204 25979: gap of unknown length
* 25980 26744: gap of unknown length
* 26745 27780: gap of unknown length
* 27781 28447: gap of unknown length
* 28448 29203: gap of unknown length
* 29204 29952: gap of unknown length
* 29953 30715: gap of unknown length
* 30716 31477: gap of unknown length
* 31478 32218: gap of unknown length
* 32219 32970: gap of unknown length

```

* 32971 33745: contig of 775 bp in length
* 33746 34540: contig of 795 bp in length
* 34541 35368: contig of 828 bp in length
* 35369 36110: contig of 742 bp in length
* 36111 36883: contig of 773 bp in length
* 36884 37645: contig of 762 bp in length
* 37646 38401: contig of 756 bp in length
* 38402 39162: contig of 761 bp in length
* 39163 40020: contig of 858 bp in length
* 40021 40780: contig of 760 bp in length
* 40781 41521: contig of 741 bp in length
* 41522 42278: contig of 757 bp in length
* 42279 43046: contig of 768 bp in length
* 43047 43829: contig of 783 bp in length
* 43830 44587: contig of 758 bp in length
* 44588 45343: contig of 756 bp in length
* 45344 46097: contig of 754 bp in length
* 46098 46850: contig of 753 bp in length
* 46851 47613: contig of 763 bp in length
* 47614 48378: contig of 765 bp in length
* 48379 49134: contig of 756 bp in length
* 49135 49897: contig of 763 bp in length
* 49898 50661: contig of 764 bp in length
* 50662 51421: contig of 760 bp in length
* 51422 52189: contig of 768 bp in length
* 52190 52956: contig of 767 bp in length
* 52957 53718: contig of 762 bp in length.

```

FEATURES

Source

```

1. 53718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"

```

```

BASE COUNT 15526 a 11338 c 11023 g 15476 t 355 others
ORIGIN

```

```

Query Match 87.0% Score 17.4: DB 43: Length 53718:
Best Local Similarity 90.0% Pred. No. 60:
Matches 18: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

```

```

Oy 1 cctgtatgacctgcaggaag 20
    |||||
Db 14737 CCTGTATGACCCACGAGAG 14718

```

```

RESULT 6
HS756G23/C LOCUS
DEFINITION
ACCESSION AL035681.13 GI:4902689
VERSION
KEYWORDS HTG: Cpg Island.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 89948)
REFERENCE
AUTHORS Corby, N.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequests@sanger.ac.uk
On May 27, 1999 this sequence version replaced gi:4775627.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM: EMBL; SW: SWISSPROT; TR: TREMBL
IMPORTANT: This sequence is not the entire insert of clone 756G23.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone dj979N1 is at 89849 in this sequence.
The true right end of clone 85F18 is at 80441 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An alt. pt is made to resolve all sequencing problems,
such as compressio. and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
feature key.
This sequence was generated from part of bacterial clone contigs of
human Chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
756G23 is from the library RPCI4 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: pcrPAC2.

```

FEATURES

source

```

1. 89948
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q13.31-13.33"
/clone="RP4-756G23"
/clone_lib="RPCI-4"
1. 247

```

```

repeat_region repeat: matches 59. 296 of consensus"
350. 571
repeat_region repeat: matches 12. 262 of consensus"
571. 662
repeat_region repeat: matches 172. 262 of consensus"
663. 948
repeat_region repeat: matches 1. 307 of consensus"
949. 1069
repeat_region repeat: matches 47. 112 of consensus"
1084. 1264
repeat_region repeat: matches 121. 309 of consensus"
1265. 1557
repeat_region repeat: matches 1. 293 of consensus"
1558. 1853
repeat_region repeat: matches 3. 300 of consensus"
1558. 1873

```

```

/note="AluJb repeat: matches 118. .121 of consensus"
repeat_region 1889. .2013
/note="MIR repeat: matches 62. .206 of consensus"
repeat_region 2172. .2466
/note="AluSg1 repeat: matches 1. .296 of consensus"
repeat_region 2533. .2832
/note="AluSx repeat: matches 1. .297 of consensus"
repeat_region 3566. .3743
/note="AluSq repeat: matches 143. .310 of consensus"
repeat_region 3744. .4049
/note="AluY repeat: matches 6. .311 of consensus"
repeat_region 4050. .4189
/note="AluSg repeat: matches 1. .143 of consensus"
repeat_region 4628. .4930
/note="AluJb repeat: matches 1. .299 of consensus"
repeat_region 4974. .5282
/note="AluY repeat: matches 1. .308 of consensus"
repeat_region 5283. .5589
/note="AluJb repeat: matches 1. .313 of consensus"
repeat_region 5731. .5815
/note="MER5B repeat: matches 20. .102 of consensus"
repeat_region 5823. .6004
/note="AluSg repeat: matches 120. .301 of consensus"
repeat_region 6005. .6039
/note="MER5B repeat: matches 11. .48 of consensus"
repeat_region 6040. .6335
/note="AluYb8 repeat: matches 1. .302 of consensus"
repeat_region 6363. .6663
/note="AluYb8 repeat: matches 1. .306 of consensus"
repeat_region 6683. .6729
/note="MIR repeat: matches 97. .145 of consensus"
repeat_region 6752. .7015
/note="AluJb repeat: matches 3. .299 of consensus"
repeat_region 7580. .7755
/note="MER5A repeat: matches 7. .184 of consensus"
repeat_region 7945. .8036
/note="MIR repeat: matches 51. .139 of consensus"
repeat_region 8094. .8156
/note="MER5B repeat: matches 108. .178 of consensus"
repeat_region 8157. .8413
/note="AluSc repeat: matches 1. .257 of consensus"
repeat_region 8414. .8555
/note="AluSc repeat: matches 1. .108 of consensus"
repeat_region 8556. .8677
/note="MER5B repeat: matches 107. .251 of consensus"
repeat_region 8735. .8878
/note="MIR repeat: matches 55. .224 of consensus"
repeat_region 9384. .9485
/note="MIR repeat: matches 86. .183 of consensus"
repeat_region 11657. .11940
/note="AluSx repeat: matches 1. .289 of consensus"
repeat_region 12040. .12351
/note="AluSg repeat: matches 1. .308 of consensus"
repeat_region 13156. .13462
/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 13493. .13626
/note="AluJb repeat: matches 1. .135 of consensus"
repeat_region 13832. .14106
/note="AluY repeat: matches 1. .304 of consensus"
repeat_region 15009. .15035
/note="L2 repeat: matches 2722. .2745 of consensus"
repeat_region 15036. .15328
/note="AluY repeat: matches 1. .296 of consensus"
repeat_region 15329. .15628
/note="AluSg repeat: matches 1. .296 of consensus"
repeat_region 15629. .15744
/note="L2 repeat: matches 2588. .2722 of consensus"
repeat_region 15928. .16054
/note="FLAMC repeat: matches 13. .133 of consensus"
repeat_region 16069. .16376
/note="AluSx repeat: matches 6. .309 of consensus"
repeat_region 17747. .17885
/note="MIR repeat: matches 48. .186 of consensus"

```

```

repeat_region 18933. .19033
/note="L2 repeat: matches 2575. .2678 of consensus"
repeat_region 19423. .19717
/note="AluSx repeat: matches 3. .297 of consensus"
repeat_region 21995. .22293
/note="AluSg repeat: matches 1. .298 of consensus"
repeat_region 22402. .22687
/note="AluJb repeat: matches 1. .288 of consensus"
repeat_region 24842. .24969
/note="L2 repeat: matches 2529. .2680 of consensus"
repeat_region 25478. .25774
/note="AluY repeat: matches 2. .298 of consensus"
repeat_region 25642. .25938
/note="L2 repeat: matches 2506. .2708 of consensus"
repeat_region 26437. .26601
/note="MIR repeat: matches 3. .173 of consensus"
repeat_region 26606. .26899
/note="AluSg repeat: matches 1. .294 of consensus"
repeat_region 26900. .27200
/note="MIR repeat: matches 173. .466 of consensus"
repeat_region 27798. .27974
/note="AluSc repeat: matches 138. .308 of consensus"
repeat_region 27975. .28278
/note="AluY repeat: matches 1. .294 of consensus"
repeat_region 28279. .28414
/note="AluSg repeat: matches 1. .138 of consensus"
repeat_region 29002. .29083
/note="MIR repeat: matches 65. .157 of consensus"
repeat_region 29080. .29296
/note="MIR repeat: matches 25. .262 of consensus"
repeat_region 29301. .29597
/note="AluSx repeat: matches 1. .297 of consensus"
repeat_region 29674. .29718
/note="MIR repeat: matches 107. .151 of consensus"
repeat_region 30167. .30254
/note="MIR repeat: matches 22. .114 of consensus"
repeat_region 30372. .30540
/note="L2 repeat: matches 2513. .2707 of consensus"
repeat_region 30943. .32264
/note="CPG island"
evidence-not-experimental
32358. .32573
/note="L2 repeat: matches 2370. .2624 of consensus"
repeat_region 32700. .32845
/note="MIR repeat: matches 1. .152 of consensus"
repeat_region 33210. .33276
/note="MIR repeat: matches 107. .171 of consensus"
repeat_region 33500. .33794
/note="AluJb repeat: matches 23. .305 of consensus"
repeat_region 34840. .35023
/note="L1M1 repeat: matches 5687. .5884 of consensus"
repeat_region 35151. .35287
/note="AluJb/FLAM repeat: matches 1. .131 of consensus"
repeat_region 35288. .35562
/note="AluY repeat: matches 1. .296 of consensus"
repeat_region 35563. .35664
/note="51 copies 2 mer ta 86% conserved"

```

Query Match 87.0% Score 17.4; DB 11; Length 89948;
Best Local Similarity 94.7% Pred. No. 63;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ctgtatgacctgcaggaag 20
DB 1038 ctgtatgacctgcaggaag 1020

RESULT 7
HS74714 106319 bp DNA PRI 23-NOV-1999
LOCUS HS74714/c
DEFINITION Homo sapiens DNA sequence from PAC 74714 on chromosome 1q23-24.
Contains a Brachyury (T box protein)-LIKE gene, ESTs, STS and a CA repeat polymorphism.

ACCESSION AL009051
VERSION AL009051.1 GI:2995195
KEYWORDS Brachyury; CA repeat; T box.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106319)
AUTHORS Wray, P.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1998) sanger.ac.uk/HGP/Chr1/ Sanger Centre,
Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 28, 1998 this sequence version replaced gi:2644985.
IMPORTANT: This sequence is the entire insert of clone 747L4.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre chromosome 1
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsur.'
feature key.
The true right end of clone 747L4 is at 1 in this sequence. The
true right end of this clone is at 106319.
747L4 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pietro de Jong.
For further details see http://bacpac.med.buffalo.edu/
Location/Qualifiers
1..106319
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q23-q24"
/clone="RP4-747L4"
/clone_lib="RPC1-4"
1..83
/note="L1MA5A repeat: matches 960. .1042 of consensus"
repeat_region
1067..1370
/note="AluS4 repeat: matches 13. .302 of consensus"
repeat_region
3124..3228
/note="MIR repeat: matches 66. .168 of consensus"
repeat_region
3236..3687
/note="L1MA7 repeat: matches 457. .920 of consensus"
repeat_region
3705..4005
/note="AluS4 repeat: matches 3. .302 of consensus"
repeat_region
5016..5049
/note="17 copies of 2 mer 94 % conserved"
repeat_region
5280..5367
/note="L1R8 repeat: matches 474. .390 of consensus"
repeat_region
5654..5730
/note="L1R8 repeat: matches 84. .1 of consensus"
repeat_region
5900..6011
/note="2 copies of 56 mer 100 % conserved"
repeat_region
8036..8199
/note="MERS5 repeat: matches 154. .2 of consensus"
repeat_region
8399..8524
/note="MIR2 repeat: matches 142. .14 of consensus"
misc_feature
8402..8900
/note="match: STS G35325"
prim_transcript
complement(8402..9600)
/note="match: ESTs N3939 W03699"
8892..9193
/note="AluJ repeat: matches 302. .3 of consensus"
gene
complement(9859..41729)

CDS
/gene="dJ747L4.1"
complement(join(<9859..10118,13949..14084,17630..17818,
22343..22404,25041..25102,29548..29682,31402..31666,
41533..41729))
/gene="dJ747L4.1"
/note="match: proteins O15178 P79778 P24781 Q25113 P79777
P20293 O02737 Q42100 O07998 P80492 Q17134 P56158 P55965;
match: cDNAs U67087 X51683 U25176 AB001871 AJ001699 D89442
S57147 M77243 X91903 D50332 S74163 D16441; match: genomic
DNA AJ001528"
/codon_start=1
/evidence=not-experimental
/product="dJ747L4.1 (Brachyury (T box protein))-like
protein"
/protein_id="CAA15624.1"
/db_xref="GI:3900891"
/db_xref="SPTREMBL:O60806"
/translation="MSLGRKSPSGTYSHLNYVESLQAGREKDPTEKQDITILE
DAPLWQREKEVINEMIYTRNGRRMFYKISTVTGLDPNMTSLDIPYDISHRKIV
NDEWPAKREVSNSHCYVIFDPSNFGAHMKAPISRYKLTKLNGGQIMLSL
HYEPDVHIVRGSARHMYTNCSPETQFIATAYONEEITALKIKYMPFAKFLDAK
ERNHLADVEALSESQHYVYSLGGMIFSNPDGCTAGNSNYQAAPLPAPPHHG
CEHYSGLRGRQAPYPSAYMHRHSVNLIESNNLOVPSGPDSTSLSTPSASI
LSYPRHNGPINCPSRPYCTMTISNAGCSPSGCEVHASPGLLGNPAVTSPSV
LSTQAPTSAQVEVLEPSLITAVSTWTAHVAHPAGMGGPGA"
10517..10722
/note="MIR repeat: matches 38. .260 of consensus"
10852..11019
/note="MERS5 repeat: matches 157. .11 of consensus"
11455..11593
/note="MIR repeat: matches 63. .202 of consensus"
11784..11995
/note="MERS3 repeat: matches 302. .82 of consensus"
12103..12303
/note="MERS3 repeat: matches 228. .1 of consensus"
12324..12433
/note="MERS5A repeat: matches 30. .163 of consensus"
12516..12806
/note="AluS4 repeat: matches 297. .1 of consensus"
12873..13117
/note="MIR repeat: matches 261. .13 of consensus"
13435..13735
/note="AluS4 repeat: matches 301. .1 of consensus"
14546..14679
/note="FLAM-C repeat: matches 1. .133 of consensus"
14705..15083
/note="L1MA8 repeat: matches 639. .1019 of consensus"
15027..15040
/gene="dJ747L4.1"
15238..15616
/note="MSTD repeat: matches 1. .394 of consensus"
16054..16110
/note="MERSA repeat: matches 91. .35 of consensus"
16180..16480
/note="AluS4 repeat: matches 303. .1 of consensus"
16490..16565
/note="MIR2 repeat: matches 60. .135 of consensus"
16817..16945
/note="MIR2 repeat: matches 4. .144 of consensus"
18036..18170
/note="L1R8 repeat: matches 1. .146 of consensus"
18360..18686
/note="L1R8 repeat: matches 310. .625 of consensus"
19428..19773
/note="L1PA5 repeat: matches 542. .890 of consensus"
21266..21557
/note="AluS4 repeat: matches 1. .292 of consensus"
23054..23348
/note="AluJ repeat: matches 2. .299 of consensus"
23609..23894
/note="AluS4 repeat: matches 3. .302 of consensus"
23876..24232
/gene="dJ747L4.1"

```

repeat_region 24617..24911
/Note="AluY repeat: matches 295..1 of consensus"
repeat_region 25217..25267
/Note="AluSx/9 repeat: matches 3..53 of consensus;
incomplete repeat"
repeat_region 25460..25555
/Note="MIR2 repeat: matches 137..37 of consensus"
repeat_region 25767..25911
/Note="MIR2 repeat: matches 2..146 of consensus"
repeat_region 25950..26099
/Note="AluSg repeat: matches 288..134 of consensus;
incomplete repeat"
repeat_region 26165..26522
/Note="THEIC repeat: matches 2..371 of consensus"
repeat_region 27570..27815
/Note="AluSg repeat: matches 296..1 of consensus"
repeat_region 28025..28325
/Note="AluSx repeat: matches 1..301 of consensus"
repeat_region 28326..28492
/Note="AluSg repeat: matches 134..300 of consensus;
incomplete repeat"
repeat_region 28497..28589
/Note="LIME2 repeat: matches 478..568 of consensus"
repeat_region 29500..29541
/Note="21 copies of 2 mer 100 & conserved"
repeat_region 29943..30242
/Note="AluY repeat: matches 301..2 of consensus"
repeat_region 32073..32270
/Note="AluJo repeat: matches 85..295 of consensus;
incomplete repeat"
repeat_region 32597..32662
/Note="LIMB6 repeat: matches 812..747 of consensus"
repeat_region 32688..32914
/Note="AluSx repeat: matches 224..1 of consensus;
incomplete repeat"
repeat_region 33089..33190
/Note="MIR repeat: matches 260..162 of consensus"
repeat_region 33838..34067
/Note="Aluub repeat: matches 73..302 of consensus;
incomplete repeat"
repeat_region 34284..34400
/Note="FLAM_C repeat: matches 131..6 of consensus"
repeat_region 34443..34518
/Note="WER2 repeat: matches 345..268 of consensus"
repeat_region 34546..34752
/Note="WER2 repeat: matches 210..1 of consensus"
repeat_region 34840..34962
/Note="WER2 repeat: matches 210..1 of consensus"
repeat_region 34845..34888
/Note="MIR repeat: matches 152..109 of consensus"

Query Match 87.0%: Score 17.4; DB 11; Length 106319;
Best Local Similarity 94.7%: Pred. No. 65;
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cctgtatgacctgcaggaa 19
|||||
DB 90330 CCTGATGAC TCGACGAA 90312

RESULT 8
AC009389 116967 bp DNA HTG 20-SEP-1999
LOCUS Drosophila melanogaster chromosome 3 clone BACR01F12 (DI1005)
DEFINITION RPCI-98 01.F.12 map 98C-98C strain Y; cn bw sp, *** SEQUENCING IN
PROCESSES ***, 93 unordered pieces.
AC009389
AC009389.5 GI:5912649
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

```

REFERENCE
AUTHORS
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 116967)
Celisner,S.E., Abgaryan,A., Arcsina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 116967)
Celisner,S.E., Abgaryan,A., Arcsina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (20-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Sep 20, 1999 this sequence version replaced g1:5832523.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bd@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 93 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 747: contig of 747 bp in length
748 827: gap of unknown length
828 1318: contig of 491 bp in length
1319 1398: gap of unknown length
1399 1986: contig of 588 bp in length
1987 2066: gap of unknown length
2067 2641: contig of 575 bp in length
2642 2722: gap of unknown length
2722 3379: contig of 558 bp in length
3379 3359: gap of unknown length
3359 3280: gap of unknown length
3280 3360: gap of unknown length
3360 3817: contig of 458 bp in length
3817 3898: gap of unknown length
3898 4897: contig of 1000 bp in length
4897 4978: gap of unknown length
4978 5629: contig of 652 bp in length
5629 5709: gap of unknown length
5709 5710: gap of unknown length
5710 6345: contig of 635 bp in length
6345 6425: gap of unknown length
6425 6934: contig of 510 bp in length
6934 7015: gap of unknown length
7015 7911: contig of 897 bp in length
7911 7991: gap of unknown length
7991 8630: contig of 639 bp in length
8630 8710: gap of unknown length
8710 9259: contig of 549 bp in length
9259 9339: gap of unknown length
9339 9340: contig of 665 bp in length
9340 10004: gap of unknown length
10004 10085: contig of 785 bp in length
10085 10870: gap of unknown length
10870 10949: gap of unknown length
10949 11505: contig of 536 bp in length
11505 11506: gap of unknown length
11506 11586: gap of unknown length
11586 12603: contig of 1018 bp in length

```

```

* 12604 12683: gap of unknown length
* 12684 13677: contig of 994 bp in length
* 13678 13757: gap of unknown length
* 13758 14648: contig of 891 bp in length
* 14649 14728: gap of unknown length
* 14729 15864: contig of 1136 bp in length
* 15865 15944: gap of unknown length
* 15945 16868: contig of 924 bp in length
* 16869 16948: gap of unknown length
* 16949 17912: contig of 964 bp in length
* 17913 17992: gap of unknown length
* 17993 18748: contig of 756 bp in length
* 18749 18828: gap of unknown length
* 18829 19541: contig of 713 bp in length
* 19542 19621: gap of unknown length
* 19622 21172: contig of 1551 bp in length
* 21173 21252: gap of unknown length
* 21253 22034: contig of 782 bp in length
* 22035 22114: gap of unknown length
* 22115 22996: contig of 882 bp in length
* 22997 23076: gap of unknown length
* 23077 24103: contig of 1107 bp in length
* 24104 24263: gap of unknown length
* 24264 24846: contig of 583 bp in length
* 24847 24926: gap of unknown length
* 24927 26566: contig of 1640 bp in length
* 26567 27522: gap of unknown length
* 27523 27602: contig of 876 bp in length
* 27603 28577: gap of unknown length
* 28578 28657: contig of 975 bp in length
* 28658 29778: gap of unknown length
* 29779 29858: contig of 1121 bp in length
* 29859 30517: gap of 659 bp in length
* 30518 30957: gap of unknown length
* 30958 31572: contig of 975 bp in length
* 31573 31652: gap of unknown length
* 31653 32385: contig of 733 bp in length
* 32386 32465: gap of unknown length
* 32466 33129: contig of 664 bp in length
* 33130 33209: gap of unknown length
* 33210 34414: contig of 1205 bp in length
* 34415 34494: gap of unknown length
* 34495 35703: contig of 1209 bp in length
* 35704 35783: gap of unknown length
* 35784 36586: contig of 873 bp in length
* 36587 36736: gap of unknown length
* 36737 37871: contig of 1135 bp in length
* 37872 37951: gap of unknown length
* 37952 39259: contig of 1308 bp in length
* 39260 39339: gap of unknown length
* 39340 39956: contig of 617 bp in length
* 39957 40036: gap of unknown length
* 40037 41349: contig of 1313 bp in length
* 41350 41429: gap of unknown length
* 41430 42775: contig of 1346 bp in length
* 42776 42855: gap of unknown length
* 42856 44226: gap of unknown length
* 44227 44306: contig of 1371 bp in length
* 44307 44942: gap of unknown length
* 44943 45022: gap of unknown length
* 45023 46435: contig of 1413 bp in length
* 46436 46515: gap of unknown length
* 46516 47405: contig of 890 bp in length
* 47406 47485: gap of unknown length
* 47486 49222: contig of 1737 bp in length
* 49223 49302: gap of unknown length
* 49303 50123: contig of 821 bp in length
* 50124 50203: gap of unknown length
* 50204 51739: contig of 1536 bp in length
* 51740 51819: gap of unknown length
* 51820 53455: contig of 1636 bp in length
* 53456 53535: gap of unknown length

```

```

* 53536 55569: contig of 2034 bp in length
* 55570 55649: gap of unknown length
* 55650 57240: contig of 1591 bp in length
* 57241 57320: gap of unknown length
* 57321 58691: contig of 1371 bp in length
* 58692 58771: gap of unknown length
* 58772 60696: contig of 1925 bp in length
* 60697 60776: gap of unknown length
* 60777 62740: contig of 1964 bp in length
* 62741 62820: gap of unknown length
* 62821 64463: contig of 1643 bp in length
* 64464 64543: gap of unknown length
* 64544 67207: contig of 2664 bp in length
* 67208 67287: gap of unknown length
* 67288 69092: contig of 1805 bp in length
* 69093 69172: gap of unknown length
* 69173 72712: contig of 3540 bp in length
* 72713 72793: gap of unknown length
* 72793 74880: contig of 2088 bp in length
* 74881 74960: gap of unknown length
* 74961 77394: contig of 2434 bp in length
* 77395 77474: gap of unknown length
* 77475 80512: contig of 3038 bp in length
* 80513 80592: gap of unknown length
* 80593 83714: contig of 3121 bp in length
* 83714 83793: gap of unknown length
* 83794 88050: contig of 4257 bp in length
* 88051 88130: gap of unknown length
* 88131 90875: contig of 2745 bp in length
* 90876 90955: gap of unknown length
* 90956 95358: contig of 4403 bp in length
* 95359 95438: gap of unknown length
* 95439 101482: contig of 6044 bp in length
* 101483 101562: gap of unknown length
* 101563 102222: contig of 660 bp in length
* 102223 102302: gap of unknown length
* 102303 102995: contig of 693 bp in length
* 102996 103075: gap of unknown length
* 103076 103600: contig of 525 bp in length
* 103601 103680: gap of unknown length
* 103681 104345: contig of 665 bp in length
* 104346 104425: gap of unknown length
* 104426 104975: contig of 550 bp in length

Query Match      87.0%: Score 17.4; DB 41; Length 116967;
Best Local Similarity 94.7%: Pred. No. 65;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 cctgtatgacctgcagaa 19
Db 19056 CCTGTATGCCCTGCAGAA 19074
||||||| |||||||

RESULT 9
AC011140 139134 bp DNA HTG 08-DEC-1999
LOCUS Homo sapiens clone RP11-116, *** SEQUENCING IN PROGRESS ***, 23
DEFINITION unorderd pieces.
ACCESSION AC011140
VERSION AC011140.2 GI:6539376
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 139134)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-116
JOURNAL Unpublished
2 (bases 1 to 139134)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslajsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

```

Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Funk, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lien, C., Locke, K., MacDonald, P., Marquis, N.,
 McEwan, P., McGuire, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Tirelli, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 Submitted (01-OCT-1999) this sequence version replaced g1:6006163.
 On Dec 8, 1999 this sequence version replaced g1:6006163.

COMMENT
 TITLE
 JOURNAL
 All repeats were identified using RepeatMasker:
 Smt, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information
 Center project name: 1_1-6

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1297: contig of 1297 bp in length
 * gap of unknown length
 * 1298 2645: contig of 1348 bp in length
 * gap of unknown length
 * 2646 4193: contig of 1548 bp in length
 * gap of unknown length
 * 4194 5882: contig of 1689 bp in length
 * gap of unknown length
 * 5883 7953: contig of 2071 bp in length
 * gap of unknown length
 * 7954 9466: contig of 1513 bp in length
 * gap of unknown length
 * 9467 12511: contig of 3045 bp in length
 * gap of unknown length
 * 12512 15375: contig of 2864 bp in length
 * gap of unknown length
 * 15376 18067: contig of 2692 bp in length
 * gap of unknown length
 * 18068 22522: contig of 4455 bp in length
 * gap of unknown length
 * 22523 28254: contig of 5732 bp in length
 * gap of unknown length
 * 28255 32322: contig of 4068 bp in length
 * gap of unknown length
 * 32323 38431: contig of 6109 bp in length
 * gap of unknown length
 * 38432 43919: contig of 5488 bp in length
 * gap of unknown length
 * 43920 51206: contig of 7287 bp in length
 * gap of unknown length
 * 51207 58769: contig of 7563 bp in length
 * gap of unknown length
 * 58770 69046: contig of 10277 bp in length
 * gap of unknown length
 * 69047 80029: contig of 10983 bp in length
 * gap of unknown length
 * 80030 88523: contig of 8494 bp in length
 * gap of unknown length

FEATURES

source
 * 88524 98308: contig of 9785 bp in length
 * gap of unknown length
 * 98309 112715: contig of 14407 bp in length
 * gap of unknown length
 * 112716 124886: contig of 12171 bp in length
 * gap of unknown length
 * 124887 139134: contig of 14248 bp in length.
 Location/Qualifiers
 1. 139134 "Homo sapiens"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="RPC1-116"
 /clone_lib="RPC1-11 Human Male BAC"
 32063 c 29598 g 38453 t 23 others

BASE COUNT

38997 a 32063 c 29598 g 38453 t 23 others
 ORIGIN

Query Match 85.0%; Score 17; DB 45; Length 139134;
 Best Local Similarity 100.0%; Pred. No. 1; e=0; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 0;

QY 1 cctgtatgacctgcag 17
 Db 64302 CCTGTATGACCTGCAG 64318

RESULT 10
 HSMYOH 3374 bp mRNA PRI 28-JUL-1995

LOCUS Human mRNA fragment for perinatal myosin heavy chain.
 DEFINITION Y00821 GI:344863
 ACCESSION Y00821 GI:344863
 VERSION myosin; myosin heavy chain.
 KEYWORDS myosin; myosin heavy chain.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 3374);
 Leinwand, L.A.;
 Direct Submission
 Submitted (12-APR-1989) Leinwand L.A., Albert Einstein College of
 Medicine, 1300 Morris Park Avenue, Bronx, N.Y. 10461, USA
 2 (bases 1 to 3374)
 Feghali, R. and Leinwand, L.A.
 Molecular genetic characterization of a developmentally regulated
 human perinatal myosin heavy chain
 J. Cell Biol. 108 (5), 1791-1797 (1989)
 89234168

JOURNAL MEDLINE
 FEATURES
 source
 Location/Qualifiers
 1. 3374
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /dev_stage="fetal"
 /tissue_type="cardiac muscle"
 /clone_id="lambda gt10"
 /clone_lib="psmHCP"
 <1. 3327
 /codon_start=1
 /product="myosin heavy chain (1078 AA)"
 /protein_id="GAA68757.1"
 /db_xref="GI:34864"
 /db_xref="SWISS-PROT:P13535"
 /translation="OKTDELAKSEAKKELEKMYLLIKEKNDLOQOVSEADSLAD
 AEERPOIKKKIOLEAKIKETVEAEEDIEINELAKKRLLEDSELSKQIDIDE
 LTLAVVEOKKATEENKVKNLTEEMAGLDDETIAKLSKEKRLQETHOOTLIDIDDEK
 QDEKLEKEEFISNLSIKIDPEOAVETLOKRIKELQARIEEGEIEARASRAKA
 EKORSDI SELEIEISERLEAGGATSAOVELINKREAFOKLRDLEATLQHEAWA
 ALRKHALSMALGEOIDNLOVKOKLEKESSEIKWEDDLSNNAEISAKGHEKMA
 CSTLEOVGGLTKTEEOQRLINDLTARARLQTEAGETSRQIDEDALVQLSKSKQ
 ASTQOIEHKKHOLEETKAKNALAHALSSRDQDLRPOVEEYBEPEGKAELOASKA
 NSSEVAWRKRIETDAORTLELEAKKILAQRLQAEHEVAVNKKASLEKTKORLO
 NEVEDIMLDEVERSNACALDKKORNFQVLSMKQVETQAELEASOKESRSLSTE

CDS

1
 1

LFKXKVNVEESLDOLETIRAKHNLQOEISDTLEOIAEGCKOIHLEKIKKOV.CEKC
EIQALEAEASLEHEBGLILRIQLELQNVSEVDRIKIAEODEITDOKRHHVET
MOSTLDAEIRSRNDALRVKRMEDLNMETOLHANNLAESLRNTFNGILIKETO
LHLDLAGEDEDLQOLAIERVERANLQAELEEMALTEOTERSKRTAEDLDASER
VOLHTONTSLINFKKLENDVSOSEVEEIOESNAEERKAKATJDAAMAEIK
KEODTSALFERKKLEJOTVAKDLOHIDEAOLKGGKKOIOLELRVRELEEVN
EOKRNEAVKGIKRRKDRVKELITOTEDRKNVLRLODILQOLAKVYSYKROAEAE
EOSNANLSKFRLOHELEAEERADIAESOVNKLKRVASREVHTAISAE"

BASE COUNT 1180 a 681 c 949 g 564 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 3374;
Best Local Similarity 90.0%; Pred. No. 94;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 450 CCTGTATGACCTCAGGAAG 20
|||||

RESULT 11
LOCUS MUSAE3A 4060 bp mRNA ROD 11-JUN-1993
DEFINITION Mouse Ael3 mRNA, complete cds.
ACCESSION M28383
VERSION M28383.1 GI:191735
KEYWORDS AEl3 protein.
SOURCE Mouse (strain BALB/c) brain, cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4060)
Kopito, R.R., Lee, B.S.M., Simmons, D.M., Lindsey, A.E., Morgans, C.W.
and Schneider, K.
Regulation of intracellular pH by a neuronal homolog of the
erythrocyte anion exchanger
Cell 59, 927-937 (1989)
Draft entry and computer-readable copy of sequence [1] kindly
submitted by R.R.Kopito, 26-SEP-1989.
Location/Qualifiers
1..4060
/organism="Mus musculus"
/db_xref="taxon:10090"
187..3870
/note="AEl3 protein"
/codon_start=1
/protein_id="AA37184.1"
/translation="MANGVTPPGGASPIPOYRVLEPEPLGPDVDEEDDLGKTLLAV
/SRTGDLISKTIPADPEKPSKSYSEDFEHRHTSHHTTHPLSARLPKPKRPPPTS
ARITRRKREKETSAPPSEGTPEIOEGGAGAEDEEEDEEESAEAPVPLPPG
POKAFPSIGSEDEDSGLPVKAPCAKALPSVGLSOSQSPQRSGSSPPPARSRISTE
KSPGSPSASVYDLREILCGSALGNPGPQRYPTDPAEQMGSGADLDMKSHRLSDN
PGRRRLVKKPSRIQGGSGSLGAILRRKKKKKLDKRRPEFVEVLEMLDLDSOE
PKHRETAARKTFEEDVEETFERMGKPHVASLSRSLLEIRTIAGAAALDLEQTLIP
GIAHLVETIVSDQIRPEDRASVYTLILKSHRPDDDSGFPENSSSVTVVIG
NHHPTPSHGPDGAVPTMADQGPAPLPHDPKPKPLHMGEGDGRKSKLLEKI
PEDAEATVAVGCVPLEOPAGAFVRLSEAVLSEVLEVPVRFLEVLGSGSHSTSD
YHELGRSIALIMSKDLFHEAAYQADRODLGASIFLDGSTVTPPEVEGHDLRVY
GAFORELRRKREDEOTKEMTTRGGAAAGKLSLEMGSEATSEDDLQRTGSVFG
GLVRYKRRPYPHPSDLROALHSGCAVAVFIYFALISPAITFGGLGEGTGLMGVS
ELIVSTAVGVLFSLGAPOLLYVSGSLVPEEAFPKRCARODELYTLGKVVWGLW
LVFVLALVAEGTFLVRIISPTQEIFAFILSLITETPFKILVYTFEHLPPRY
PDEALFTGLNLSALPTEGPPGNOPNTALSLILMGLTFLAEFLKRRNSRFL
GGAARRIIGDGIPIILVAVLVDSYITDQIKLVPTGLSVTSPHKKTWFIPPLGS
AREPPMMVAAPALVILIFEMETQIALIVSOKARLLKSGSFGHLLDGLSGLG
GICIGLIPWLTAAVRSVTHVNAIVMTAIPADGKPOIOEYREOGRVTVGLIASVG
LSIVKGVAVRIPRLAVIFGFIYMGVTSI..310SORLLLEMPAKHHPDEOPVTVK
TWMLDFTICIQGLALMLVYKSTASLAPFLILLTLVPSGLLPRLDQDEQLQALD
SEAEPPNFDDGGDEVELHMPV"

BASE COUNT 766 a 1223 c 1220 g 851 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 4060;
Best Local Similarity 90.0%; Pred. No. 96;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 3862 CCTGTATGACCTCAGGAAG 20
|||||

RESULT 12
LOCUS RNU35022/c 4350 bp mRNA ROD 19-APR-1999
DEFINITION Rattus norvegicus cis-Golgi matrix protein GM130 mRNA, complete
cds.
ACCESSION U35022
VERSION U35022.2 GI:4583536
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 4350)
Nakamura, N., Rabouille, C., Watson, R., Nilsson, T., Hul, N.,
Slusarewicz, P., Kreis, T.E. and Warren, G.
Characterization of a cis-Golgi matrix protein, GM130
J. Cell Biol. 131 (6 Pt 2), 1715-1726 (1995)
96134022
2 (bases 1 to 4350)
Nakamura, N., Lowe, M., Levine, T.P., Rabouille, C. and Warren, G.
The vesicle docking protein p15 binds GM130, a cis-Golgi matrix
protein, in a mitotically regulated manner
Cell 89 (3), 445-455 (1997)
97294388
3 (bases 1 to 4350)
Lowe, M., Rabouille, C., Nakamura, N., Watson, R., Jackman, M.,
Jamsa, E., Rahman, D., Pappin, D.J. and Warren, G.
Glc2 kinase directly phosphorylates the cis-Golgi matrix protein
GM130 and is required for Golgi fragmentation in mitosis
Cell 94 (6), 785-793 (1998)
98424247
4 (bases 1 to 4350)
Nakamura, N.
Direct Submission
Submitted (29-AUG-1995) Nobuhiro Nakamura, Cell Biology Laboratory,
Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London,
WC2A 3PX, UK
5 (bases 1 to 4350)
Nakamura, N.
Direct Submission
Submitted (14-APR-1999) Department of Molecular Biology, Kyushu
University, Fukuoka 812-8582, Japan
Sequence update by submitter
On Apr 14, 1999 this sequence version replaced gi:2062177.
Location/Qualifiers
1..4350
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
85..3045
/codon_start=1
/product="cis-Golgi matrix protein GM130"
/protein_id="AAB53335.2"
/db_xref="GI:4583537"

REMARK
COMMENT
FEATURES
source

CDS

BASE COUNT 766 a 1223 c 1220 g 851 t

```

AMMOORVOOMAEOYHALKEEKREORESOVLESLAELRSOMEPEPPPEPTGSEAE
ERLGEVEFOLQKLEGLTGLOLRAOVONDSLSHNRQEGMLLELREDAQHSQDAEE
KROTESQSOPTTISRALSQNRREKEALONGPVRLTNENNETSALQSEHYK
ELARLGLQERLGLKTEVELKESQAGLOEQDQCSHLOQYAAATQOHLVYEDL
TSEKALHKOILLQLOMDLOLQHEVOGKMAEMARQELOERLAKASSQENQLOLA
OLSLVLQGEQMDQDEQDEVPQPSLIPEDLVSRAMAVAFQNAIARAEEQOARLV
OLREKACRSIAHLAIPVOSKLEAVVPDMGDSVSEENQALHYAMEKLOSRFLE
VMOEKVELKEVELEHCCIOISGETDITGEI LYONQRAVLKARHLEKEEYISRLA
ODKEBMKYLELOELVRLVNERENGRKLAVSONPRAVPVPQSGOEFCAADQO
GDLREVSADDTPEPAQGEAGVPAPQENPTAQOITMQLREIQNPQERFGLGSPCTPF
YRABNDQEVKIMVI"
misc-feature
/site="phosphorylated by cdc2 kinase; phosphorylation
157. 159
Base COUNT 1121 a 1109 c 1312 g 808 t
ORIGIN
Query Match 84.0% Score 16.8; DB 12; Length 4350;
Best Local Similarity 90.0% Pred. No. 96;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 cctgtatgacctgcaggag 20
11111111111111111111
Db 2488 CCTGCATGACCTCCAGGAG 2469
RESULT 13
HSDMHC 4437 bp mRNA PRI 19-JUL-1995
LOCUS Human MHC for fetal-myosin heavy chain clone gtmHC-F.
DEFINITION X51592
VERSION X51592.1 GI:29465
KEYWORDS fetal-myosin heavy chain.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4437)
AUTHORS Robert E.
JOURNAL Direct Submission
TITLE Submitted (02-JAN-1989) Robert E., Department of Toxicology, Medical
SCHOOL, University of Hamburg, Grindelallee 117, D-2000 Hamburg 13,
FRG
REFERENCE 2 (bases 1 to 4437)
AUTHORS Arnold H.H., Buchberger-Seidl A., Braun T., Singh S., Goede H.W. and
Arnold H.H.
TITLE Identification of three developmentally controlled isoforms of
human myosin heavy chains
JOURNAL Eur. J. Biochem. 189 (1), 55-65 (1990)
MEDLINE 90235862
REFERENCE 3 (bases 1 to 4437)
AUTHORS Robert E., Lyons G.E., Braun T., Cossu G., Buckingham M. and
Arnold H.H.
TITLE The muscle regulatory gene, Myf-6, has a biphasic pattern of
expression during early mouse development
JOURNAL J. Cell Biol. 113 (6), 1255-1265 (1991)
MEDLINE 91258395
COMMENT clone-gtmHC-F.
FEATURES
source Location/Qualifiers
1..4437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="22 weeks of gestation"
/tissue_type="skeletal muscle"
/clone_1fb="lambda gtl1"
<1..4314
/codon_start=1
/product="fetal-myosin heavy chain (4437 AA)"
/protein_id="CAA35941.1"
/db_xref="GI:29466"
/db_xref="SWISS-PROT:P13535"
/translation="EYKKEGIEWTFIDFMDLACILIEKPLGIFSILEECPMPK

```

```

ADTSEKKNLYDOHLGKSNFQPKVYKGAENHSLIHVAGTVDNITGMKDKDP
LNDTVAGQKAMKTLASLSTFYSAEADSSAKGKAGSSRQTVSLFRNLNG
MTNLSGTHHPRCITIPNETKRGAMEHVLVHQRGCVGTRIGCRKSPRLNG
DKQRKRVNANSAITEGQETDKKASELNLSDIDHOYFGHRTKVFYFAGLGLLE
ENRDEKLAQITRTQAVCRGLFMRVEYQKMLQRRALFCIOYVRAEYNVNHMPML
FKRIPRLASATEKEMATKEEFOKTDKMLKSAKKEEKKVTLKAKRLEDOLO
VQSEADSLADAEERCEOLIKNKIQLEAKIKETVFAEEEEEELTAKSEKALQETFOOT
ELKQIDDELTLAKVEKKNATENKVNLTPEMGLDETAKSEKALQETFOOT
LDDLOAEEDKVNILTKAKTKLEQVQDLEGSLEQKTRMLERKRLLEDGLKAOE
STMDNDNQOJDEKLEKEFTISLISKTEDEQAVETOLOKRIEEDARILEGEEL
EAEASRAAEKORSLSLELEISERELEGAGSAVLELNKREAFQRLKRDLE
ATLOHEAMVYALRKHADSMALGEOINLQVOKLEKSELMETDLSNAEL
SKAGNLEKVGSRLEDOVELKTEEEQRLINDLTQRRARLOTEAEYSRODEG
ALVQSLRSKQASTOOIEELKHQLEETEKAKALAHALQSSRHCDLLEQVEEEOG
KAEIOLRSKASEVQWMTNTETALQRTETELEAKKKLAORLOEAEHVEAVNAK
ASLEKTKORLQNEVEDLMDVRSNAACALDKKRNFDYLSKKOKEEETOAEFA
SQRESRSLSTELFKVKNVYESLDQLETLREKNLQOESDLEQIAGEKQIHELE
KIKQVEQKCEIOALEEAEASLEHEGKILRQLELNVKSEVDRIAEKDEIDQ
LRKNHRTVETMOSTLDAEIRSDNALRVKMKMEGDLNEMETOLNHNARLAESELRN
RMTQGLKETOLDLALRGQEDLKEQALIVERANLQAEIEELMATLEQTERSRI
AEQELDASERVQULHTQNTSLINTKKLENDVQLOSEVEVEIOESNAEKKAKAI
TPAAMWAEELKKEQDPTSAHLEPRKKNLQDYVQDLOHRLDPAQALQKQKQIOLEA
RVRELEGEYENQKRNAEAVKGLRKHERVAKELITQTEDEKKNVRLQDLYDKQAVY
KSKRQAEAEEOSSANLSKFRKLQHELEAEERADIAESOVNKLRYASREVHTKISA
E"
BASE COUNT 1498 a 903 c 1214 g 822 t
ORIGIN
Query Match 84.0% Score 16.8; DB 9; Length 4437;
Best Local Similarity 90.0% Pred. No. 96;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 cctgtatgacctgcaggag 20
11111111111111111111
Db 1524 CCTGCATGACCTCCAGGAG 1543
RESULT 14
HSDMHC 5259 bp mRNA PRI 27-APR-1993
LOCUS Human desmoplakin mRNA, 3' end.
DEFINITION J05211
VERSION J05211.1 GI:181607
KEYWORDS desmoplakin.
SOURCE Human foreskin keratinocyte, cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 5259)
AUTHORS Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A.,
Wagner R.M., Angst B.D., Nilles L. and Nilles L.A.
TITLE structure of the human desmoplakins: implications for function in
the desmosomal plaque
JOURNAL J. Biol. Chem. 265, 2603-2612 (1990)
MEDLINE 90153880
REFERENCE 2 (sites)
AUTHORS Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A.,
Wagner R.M., Angst B.D., Nilles L. and Angst B.D.
TITLE Additions and Corrections: Structure of the human desmoplakins:
implications for function in the desmosomal plaque
JOURNAL J. Biol. Chem. 265, 11406-11407 (1990)
MEDLINE 90361712
COMMENT [2] revises [1].
Draft entry and computer-readable sequence for [1] kindly submitted
by K.J.Green, 12-JAN-1990.
FEATURES
source Location/Qualifiers
1..5259
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..5259
/codon_start=1
/translation="EYKKEGIEWTFIDFMDLACILIEKPLGIFSILEECPMPK

```


MDLINE 95203869
GenBank staff at the National Library of Medicine created this entry [NCBI gidsq 166051] from the original journal article.

REMARK This sequence comes from Fig. 2B.
Map location: 19D1.

FEATURES
source location/Qualifiers

1..1742
/organism="Mus sp."
/db_xref="taxon:10095"
order(S77710.1:330..654,1..1502)
/gene="Ifi54"
/note="interferon-alpha regulated gene"
join(S77710.1:330..334,89..1502)
/gene="Ifi54"
/note="interferon-alpha regulated gene; This sequence comes from Fig. 2A"

CDs
/protein_id="AAB3830.1"
/db_xref="GI:998615"
/translation="MSTTSKESLVNCLROLKCHFTWNLIAEDESLEDFEDRVENKDEF
ONSEFKATMNCNIIAYVKHCRGNEALOCLEAGFIOOHPPOVEIRSLVTGNAM
VTHMGPSKQATLDKVKYCKRFSPIRENPAIDCEGARIKCTKQNRVVC
FOKALEKDPKAPFTSGMAIAFTYLDWPARNTCIDLEQATQISPDNTTYKVLALK
LDAVHVHKNOMALVEBALKDPASIDITLLAAFTCYKYVTDRAIQLEKALELPN
NAVYVYGGCCYRSKXVHMLNREMFSGDKLEEDIQLVNHLRKAEMIKEMLEYS
CSFLADYIIAKKYDEADYFQKELSKDPPGKQILHLRYGNFOFORDKALYH
YMEGVKIKKTKTPKKMKRRIQRIALRLHEDESDSALHILAI QENGGQQAOKDS
FRGVDSANVPASALDEGAEY"

BASE COUNT 525 a 356 c 440 g 421 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 1742;
Best Local Similarity 90.0%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccagttctcaaggagcagc 20
|||||
DB 407 CCAGTCTCAAGGCTCAGG 426

RESULT 12
LOCUS MmplzFgen 2551 bp mRNA ROD 15-MAR-1995
DEFINITION M.musculus Plzf gene.
ACCESSION 247205
VERSION 247205.1 GI:732822
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
AUTHORS
JOURNAL
TITLE
DIRECT SUBMISSION
SUBMITTED (04-JAN-1995) Zelen A., Institute of Cancer Research,
Leukemia Research Fund Centre, 237 Fulham Road, London, UK, SW3
6JB
2 (bases 1 to 2551)
Zelen A.
Proc. Natl. Acad. Sci. U.S.A. (1995) In press
Expression of the zinc-finger gene PLZF at rhombomere boundaries in
the vertebrate hindbrain
Proc. Natl. Acad. Sci. U.S.A. (1995) In press

FEATURES
source location/Qualifiers

1..2551
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="mplzf"
/dev_stage="adult"
/tissue_type="heart"

BASE COUNT 622 a 717 c 761 g 451 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 2551;
Best Local Similarity 90.0%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccagttctcaaggagcagc 20
|||||
DB 1692 CCAGTCTCAAGGAGCAGG 1711

RESULT 13
LOCUS CHTR0501 2571 bp DNA VRT 28-APR-1993
DEFINITION Chicken tropomyosin beta subunit gene, exons 1-2.
ACCESSION M21223 J04476
VERSION M21223.1 GI:212822
KEYWORDS
SEGMENT
1 of 4
alternative splicing; tropomyosin.
Chicken (5-day-old embryo) erythroblast DNA, clones 2.333 and
8.631.

ORGANISM
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 2571)
Libri D., Lemonnier M., Meinel T., and Fisman M.Y.
A single gene codes for the beta subunits of smooth and skeletal
muscle tropomyosin in the chicken
J. Biol. Chem. 264, 2935-2944 (1989)
JOURNAL
AUTHORS
MEDLINE
COMMENT
Draft entry and computer-readable sequence for [1] kindly provided
by D. Libri, 12-MAR-1989.

FEATURES
source location/Qualifiers

1..2571
/organism="Gallus gallus"
/db_xref="taxon:9031"
prim_transcript 801..>2571
/note="TROSS BS mRNA and introns"
exon
/number=1
999..1700
/number=1
1701..1826
/number=2
exon
543 a 766 c 787 g 464 t 11 others
BASE COUNT 543 a 766 c 787 g 464 t 11 others
ORIGIN 5 bp upstream of PstI site.

Query Match 84.0%; Score 16.8; DB 4; Length 2571;
Best Local Similarity 90.0%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccagttctcaaggagcagc 20
|||||
DB 121 CCAGCTCTGAAGGAGCAGG 102

RESULT 14
LOCUS MMU43085 3513 bp mRNA ROD 02-JUL-1996
DEFINITION Mus musculus glucocorticoid-attenuated response gene 39 (GAR-39)
ACCESSION U43085
VERSION U43085.1 GI:1401061
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE 1 (bases 1 to 3513)
AUTHORS Smith,J.B. and Herschman,H.R.
TITLE The glucocorticoid attenuated response genes GARG-16, GARG-39, and
GARG-49/ING2 encode inducible proteins containing multiple
tetraatricopeptide repeat domains
JOURNAL Arch. Biochem. Biophys. 330 (2), 290-300 (1996)
MEDLINE 96239145
REFERENCE 2 (bases 1 to 3513)
AUTHORS Smith,J.B. and Herschman,H.R.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1995) Jeffrey B. Smith, Pediatrics, UCLA School
of Medicine, 10833 Le Conte Avenue, Los Angeles, CA 90095-1752, USA
FEATURES
Source Location/Qualifiers
1..3513
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="Swiss 3T3"
52..1464
/gene="GARG-39"
52..1464
/gene="GARG-39"
/standard_name="glucocorticoid-attenuated response gene
39"
/note="IFIT-2: contains ten tetraatricopeptide repeat (TPR)
domains: homolog of human ISG-54K"
/codon_start=1
/product="GARG-39"
/protein_id="AAC52626.1"
/db_xref="GI:1401062"
/translat="MSTTSKESLESNLRQKCHPTWNLIAEDESLEDFEDRVFNKDEF
ONSEKATMCNCLAYKHCRCRGLEALQCLGEAEGLIOOHDPDVEIESTLVTGNVAV
VYHMGOSKQAVLYDKVKQCKKSSPYRIENPALDCEEGMARLKCTKNONEVAKYC
FOKALEKPKNPPEPTSGAIAINYRIIDMPAR JYCDSDLEQAIOLSPNTYKVLTAIK
LDVHKRQNMALVEARLKKDSALDITLLRARFYCKYVDTPRAIOLLRKALEKIPNNA
VYHYVGGCCYRSKRVHMLNRREWVPSGDRKKLEELIQLVANHLKRAEIKEMLYSSS
FLADLYIAKKYDEADYVYFQELSKDLPPGKQLHLRYGNFQFNKRQOKAIYHYM
EGVKIKKKTIIPQKRRKRLORIALRLRHEDESDSEALHILAFIQENGGOADQDSER
CVDSANOVPSASLDEGAAY"
203..304
/gene="GARG-39"
/note="encodes TPR 1 domain"
312..433
/gene="GARG-39"
/note="encodes TPR 2 domain"
521..565
/gene="GARG-39"
/note="encodes TPR 3b domain"
575..676
/gene="GARG-39"
/note="encodes TPR 4 domain"
677..775
/gene="GARG-39"
/note="encodes TPR 5 domain"
776..877
/gene="GARG-39"
/note="encodes TPR 6 domain"
878..910
/gene="GARG-39"
/note="encodes TPR 7a domain"
995..1039
/gene="GARG-39"
/note="encodes TPR 7b domain"
1040..1141
/gene="GARG-39"
/note="encodes TPR 8 domain"
1151..1255
/gene="GARG-39"
/note="encodes TPR 9 domain"
1328..1426
/gene="GARG-39"
/note="encodes TPR 10 domain"
misc_count 1130 a 641 c 792 g 950 t
ORIGIN

```

```

Query Match 84.0%; Score 16.8; DB 12; Length 3513;
Best Local Similarity 90.0%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 ccaattctcaagagagcaag 20
Db 375 CCAAGTCTCAAGGCTCAGG 394
|||||
RESULT 15
AC008293
LOCUS
DEFINITION
AC008293 55741 bp DNA HTG 02-AUG-1999
Drosophila melanogaster chromosome 3 clone DS01181 (0512) map
89D-89D strain y: cn bw sp, *** SEQUENCING IN PROGRESS ***, 40
unordered pieces.
AC008293 GI:5670673
HTG: HTGS_PHASE1.
SOURCE
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 55741)
Celisner,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomoclan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paley,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequerra,A., Sethi,H., Shif,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 55741)
Celisner,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomoclan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paley,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequerra,A., Sethi,H., Shif,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases. Pl library location:
13-29.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1095: contig of 1095 bp in length
* 1096 1175: gap of unknown length
* 1176 2015: contig of 840 bp in length
* 2016 2095: gap of unknown length
* 2096 2737: contig of 642 bp in length
* 2738 2817: gap of unknown length
* 2818 4077: contig of 1259 bp in length
* 4077 4156: gap of unknown length
* 4157 4991: contig of 835 bp in length

```

```
* 4992 5071: gap of unknown length
* 5072 5668: contig of 597 bp in length
* 5669 5748: gap of unknown length
* 5749 6979: contig of 1231 bp in length
* 6980 7059: gap of unknown length
* 7060 7950: contig of 891 bp in length
* 7951 8030: gap of unknown length
* 8031 8745: contig of 715 bp in length
* 8746 8825: gap of unknown length
* 8826 9779: contig of 954 bp in length
* 9780 9860: gap of unknown length
* 9860 10832: contig of 973 bp in length
* 10833 10912: gap of unknown length
* 10913 12139: contig of 1227 bp in length
* 12140 12219: gap of unknown length
* 12220 13632: contig of 1413 bp in length
* 13633 13712: gap of unknown length
* 13713 15303: contig of 1591 bp in length
* 15304 15383: gap of unknown length
* 15384 16385: contig of 1002 bp in length
* 16386 16465: gap of unknown length
* 16466 17393: contig of 928 bp in length
* 17394 17473: gap of unknown length
* 17474 18788: contig of 1315 bp in length
* 18789 18868: gap of unknown length
* 18869 20618: contig of 1750 bp in length
* 20619 20698: gap of unknown length
* 20699 21896: contig of 1198 bp in length
* 21897 21976: gap of unknown length
* 21977 23559: contig of 1583 bp in length
* 23560 23639: gap of unknown length
* 23640 25202: contig of 1563 bp in length
* 25203 25282: gap of unknown length
* 25283 26948: contig of 1666 bp in length
* 26949 27028: gap of unknown length
* 27029 28616: contig of 1588 bp in length
* 28617 28696: gap of unknown length
* 28697 30859: contig of 2163 bp in length
* 30860 30939: gap of unknown length
* 30940 32463: contig of 1524 bp in length
* 32464 32543: gap of unknown length
* 32544 34287: contig of 1744 bp in length
* 34288 34367: gap of unknown length
* 34368 37264: contig of 2897 bp in length
* 37265 37344: gap of unknown length
* 37345 40346: contig of 3002 bp in length
* 40347 40426: gap of unknown length
* 40427 45018: contig of 4592 bp in length
* 45019 45098: gap of unknown length
* 45099 48778: contig of 3680 bp in length
* 48779 48858: gap of unknown length
* 48859 49532: contig of 694 bp in length
* 49533 49612: gap of unknown length
* 49613 50244: contig of 612 bp in length
* 50245 50324: gap of unknown length
* 50325 51056: contig of 732 bp in length
* 51057 51135: gap of unknown length
* 51137 51792: contig of 656 bp in length
* 51793 51872: gap of unknown length
* 51873 52214: contig of 342 bp in length
* 52215 52294: gap of unknown length
* 52295 52955: contig of 661 bp in length
* 52956 53035: gap of unknown length
* 53036 53673: contig of 638 bp in length
* 53674 53753: gap of unknown length
* 53754 54282: contig of 529 bp in length
* 54283 54362: gap of unknown length
* 54363 55053: contig of 691 bp in length
* 55054 55133: gap of unknown length
* 55134 55741: contig of 608 bp in length.
```

FEATURES

Source

```
Location/Qualifiers
1..55741
/organism="Drosophila melanogaster"
```

```
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"
/map="89D-89D"
/clone="DS01181 (D512)"
/clone.lib="PI library, partial Sau3A in pNS582vec14Ad10"

BASE COUNT      14707 a 11367 c 11512 g 10032 t 3123 others

ORIGIN

Query Match      84.0%  Score 16.8;  DB 33:  Length 55741;
Best Local Similarity 90.0%;  Pred. No. 85;
Matches 18;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

QY      1 ccagttctcaagagcagc 20
      ||| ||||| ||||| |||
Db 20591 CCATTCTCAAGACCGG 20610
```

```
Search completed: April 25, 2000, 20:38:36
Job time: 21124 sec
```

LOCUS HSDJ468N4 237415 bp DNA HTG 23-NOV-1999
DEFINITION Homo sapiens chromosome 20 clone RP3-468N4, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL121752
VERSION AL121752.2 GI:6136980
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Metazoa: Chordata; Craniata: Vertebrata; Mammalia:
Eutheria: Primates: Catarrhini; Homnidae: Homo.
1 (bases 1 to 237415)
REFERENCE 1
AUTHORS Steward, C.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerey@sanger.ac.uk
On Oct 27, 1999 this sequence version replaced g1:6014391.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments. Unfinished: d468N4 Contig.ID: 01553 acc=AL121752
Length: 1449 bp Unfinished: d468N4 Contig.ID: 01594
Length: 36096 bp Unfinished: d468N4 Contig.ID: 01594
acc=AL121752 Length: 35200 bp Unfinished: d468N4 Contig.ID:
01633 acc=AL121752 Length: 107601 bp Unfinished: d468N4
Contig.ID: 01678 acc=AL121752 Length: 39960 bp Unfinished:
d468N4 Contig.ID: 01724 acc=AL121752 Length: 13109 bp.
NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source 1..237415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP3-468N4"
/clone_lib="RPC1-3"
BASE COUNT 65944 a 53184 c 52924 g 61362 t 4001 others
ORIGIN

Query Match 85.0% Score 17: DB 32: Length 237415;
Best Local Similarity 100.0%: Pred .. 97;
Matches 17: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 4 gttctcaagagcagc 20
|||||
Db 102945 GTTCTCAAGAGCAGC 102929

RESULT 11
S77710S2 1742 bp DNA ROD 27-SEP-1995
LOCUS
DEFINITION If154-interferon-alpha regulated gene [mice, Genomic, 1742 nt,
segment 2 of 2].
S77713
ACCESSION S77713.1 GI:998614
VERSION
KEYWORDS
SEGMENT
SOURCE
ORGANISM
2 of 2
Mus sp.
Mus sp.
Eukaryota: Metazoa: Chordata; Craniata: Vertebrata; Mammalia:
Eutheria: Rodentia: Sciurognathi; Muridae: Murinae: Mus.
REFERENCE 1 (bases 1 to 1742)
AUTHORS Blyss, H.A., Vlietstra, R.J., Faber, P.W., Smit, E.M., Hagemeijer, A.
and Trapman, J.
TITLE Structure, chromosome localization, and regulation of expression of
the interferon-regulated mouse If154/If156 gene family
JOURNAL Genomics 24 (1), 137-148 (1994)

LOCUS HSDJ61404/c 23-NOV-1999
DEFINITION Homo sapiens chromosome 20 clone RP4-61404 map q11.1-12, ***
SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION AL121753
VERSION AL121753.3 GI:6468372
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Metazoa: Chordata; Craniata: Vertebrata; Mammalia:
Eutheria: Primates: Catarrhini; Homnidae: Homo.
1 (bases 1 to 153409)
REFERENCE 1
AUTHORS Collier, R.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerey@sanger.ac.uk
On Nov 26, 1999 this sequence version replaced g1:6224551.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments. Contig.ID: 01079 Length: 36096bp
Contig.ID: 01394 Length: 114791bp
Contig.ID: 02833 Length: 9220p.
NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source 1..153409
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q11.1-12"
/clone="RP4-61404"
/clone_lib="RPC1-4"
BASE COUNT 41740 a 35767 c 35623 g 38679 t 1400 others
ORIGIN

Query Match 85.0% Score 17: DB 32: Length 153409;
Best Local Similarity 100.0%: Pred No. 67;
Matches 17: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 4 gttctcaagagcagc 20
|||||
Db 140082 GTTCTCAAGAGCAGC 140066

RESULT 10
HSDJ468N4/c

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:38:17 ; Search time 7810.03 seconds
(without alignments)
-7.776 Million cell updates/sec

Title: US-08-956-991-5

Perfect score: 20
Sequence: 1 ccaagttccaaagagcagcag 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da1: *
2: gb_da2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_sts: *
14: gb_sy: *
15: gb_un: *
16: gb_v1: *
17: em_fun: *
18: em_huml: *
19: em_hum2: *
20: em_in: *
21: em_om: *
22: em_or: *
23: em_ov: *
24: em_pat: *
25: em_ph: *
26: em_pl: *
27: em_ro: *
28: em_sts: *
29: em_sy: *
30: em_un: *
31: em_v1: *
32: gb_htg1: *
33: gb_htg2: *
34: gb_in1: *
35: gb_in2: *
36: em_da1: *
37: em_da2: *
38: em_hum3: *
39: em_hum4: *
40: gb_pr4: *
41: gb_htg3: *
42: gb_htg4: *
43: gb_htg5: *
44: gb_htg6: *

45: gb_htg7: *
46: em_htg1: *
47: em_htg2: *
48: em_htg3: *
49: em_hum5: *
50: gb_pl3: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	6110	11 AF023449	AF023449 Homo sapi
2	20	100.0	6413	11 AF023450	AF023450 Homo sapi
3	20	100.0	145861	11 AF064862	AF064862 Homo sapi
4	20	100.0	159424	11 AF064865	AF064865 Homo sapi
5	17.4	87.0	3654	4 CHKPI05A	M86930 Chicken p10
6	17.4	87.0	3846	4 CGAF000241	AF0000241 Gallus ga
7	17.4	87.0	4023	4 CHKNERB1	D13719 Chicken mRN
8	17.4	87.0	113575	33 AL133354	AL133354 Homo sapi
9	17.4	85.0	153409	32 HSDJ61404	AL121753 Homo sapi
10	17	85.0	237415	32 HSDJ468N4	AL121752 Homo sapi
11	16.8	84.0	1742	12 S7771052	S77713 Ifi54-Inter
12	16.8	84.0	2551	12 MMRP2FCGN	247205 M.musculus
13	16.8	84.0	2571	12 CHKTR0501	M21223 Chicken tro
14	16.8	84.0	3513	12 MM043085	U43085 Mus musculu
15	16.8	84.0	55741	33 AC008293	AC008293 Drosophila
16	16.8	84.0	55795	11 AF000573	AF000573 Homo sapi
17	16.8	84.0	56559	33 AC008294	AC008294 Drosophila
18	16.8	84.0	76652	44 AC016791	AC016791 Mus muscu
19	16.8	84.0	89001	41 AC011168	AC011168 Homo sapi
20	16.8	84.0	123039	45 AC007942	AC007942 Mus muscu
21	16.8	84.0	141891	41 AC009289	AC009289 Homo sapi
22	16.8	84.0	169926	44 AC013691	AC013691 Homo sapi
23	16.4	82.0	2148	13 G15848	G15848 human SYS C
24	16.4	82.0	3188	2 RSU83438	UB3438 Rickettsia
25	16.4	82.0	12446	9 HSDMDR	X14298 Human mRNA
26	16.4	82.0	13957	9 HSDMDIS	M18533 Homo sapien
27	16.4	82.0	13957	15 M20250	M20250 Figure 1. N
28	16.4	82.0	36083	43 AC013624	AC013624 Homo sapi
29	16.4	82.0	65201	45 AC004059	AC004059 Homo sapi
30	16.4	82.0	92053	45 AC017270	AC017270 Drosophila
31	16.4	82.0	98056	40 AC006061	AC006061 Homo sapi
32	16.4	82.0	153568	11 AC002981	AC002981 Homo sapi
33	16.4	82.0	188948	40 AC005832	AC005832 Homo sapi
34	16.4	82.0	194000	33 AC000016	AC000016 Homo sapi
35	16.4	82.0	328651	42 AC010547	AC010547 Homo sapi
36	16	80.0	51648	43 AC015434	AC015434 Drosophila
37	16	80.0	78643	35 AC004352	AC004352 Drosophila
38	16	80.0	140692	44 AC013362	AC013362 Homo sapi
39	16	80.0	143712	11 HS917N8	AL031684 Human DNA
40	16	80.0	158301	42 AC009589	AC009589 Homo sapi
41	16	80.0	159469	42 AC012274	AC012274 Homo sapi
42	16	80.0	161598	45 AC016801	AC016801 Homo sapi
43	16	80.0	189565	32 CNS01D04	AL133163 Homo sapi
44	15.8	79.0	329	12 AF044224	AF044224 Mus muscu
45	15.8	79.0	370	13 A0028021	A0028021 Rattus no

ALIGNMENTS

RESULT 1
LOCUS AF023449 6110 bp mRNA
DEFINITION Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)
ACCESSION AF023449
VERSION AF023449.1 GI:3169765

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 Eutheria; Primates; Catarrhini; Homiinae; Homo.
AUTHORS 1 (bases 1 to 6110)
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,
Lyons, G.E. and Korenberg, J.R.
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a
Down Syndrome Region and is Involved in the Development of the
Nervous System
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6110)
AUTHORS Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,
Lyons, G.E. and Korenberg, J.R.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los
Angeles, CA 90048-1869, USA
FEATURES
SOURCE Location/Qualifiers
1. 6110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21 (trisomy 21)"
/map="21q22, between HMG14 and MX1"
/tissue_type="brain"
/dev_stage="14 weeks, fetal"
/clone="CHD2-42"
/note="derived from alternately-spliced mRNA"
<1. 6110
/gene="DSCAM"
<1. 5691
/gene="DSCAM"
/note="member of immunoglobulin superfamily; involved in
nervous system development"
/codon_start=1
/product="Down syndrome cell adhesion molecule"
/protein_id="AAC17966.1"
/db_xref="GI:3169766"
/translation="VSESDHSSLYPVNASLQEVFASTTGTVPCPAGIPVYLRM
YLATGEITVPGIRHVPNGTLQIFPPSPSTLIHDTYICLAEVNSGKIRSDV
HKAVALREPTVREDDKTRGNVAFAKCIIPSSVAYIIVSWEMKDYSALSGSRL
ITSTGALYIKDVONEDGLYNYRCITRHRVYGETROSNASALFVSDPANSPALDDF
HRRAMAGORVELPCKALGHEPDYRMLKDMPLLSGRFOKTVTGILINIRSDGS
YVCEVSNRYGTAVIGRLYKOPILATISPRKXSSGVSVLSGCVTDEOELSMY
RNGEILNPKGNVITGIIHNLIMDHVKSDBGAYOCFVRKDKLSQDYVOYVLEGT
PKLISAFSEKVVSPAEFVSLMCNVKGPILPTITWTDDEPILKSGSHRISOMTSEGN
VSYLNISSQVBDGYYRGTANNAGVLYQARINVRPASIRPKNTITAGRDY
IHCRVIGPYYSIKWYKNSNLPEFNHROVALFENNGLKLSVQKEVEDEEYCNVLYO
POLSTOSVHTVAKVPFIOPFEFEPFISIGORVFIQVYVSGDLPTITWOKGRIIP
GSLGVTIDNIDFTSLRLSNLIMHNGNNTCIRARNAVAEHOSOLIVAPPEFVOP
RDOGGYIKAVILNCSAEQYVPTIYVRFKSGVGFQFOPIALNGRIQVLSNGLIK
HVEEDSGITLCKVSNVGVADVSKSMILYIKIIPAMITSTPNTLTAIGOKKESCTAH
GEPTIYVWEKEDRIINPEMARLVSTKEGEVISTLQILPVREDSEFSCAHNS
YGEDRGIIQLTVEPDPEIEIKDKAKRITLMTMGEDGNPITGYDIECKNSDS
WDSKORTDVPOLNSATLIDHVSSTYSIRMAKNRIGSEPSNELITTADEAAGD
POEVHLPISSOSIRYTKAPKKHLONGIIRGOYIGREYSGNGNENITLSDVSG
DSFVYTLDNKSTFOYGLVYACNAGCTGSSGSEITITTLDEVPSPENVAALASP
ESISISWTSKELANGILOGFRTYIRANIMODELEIKNITTQPSLEDGIEKLTN
YSIOVLFTRAGGVRESEQLFTRKEDVPEPAGVAAAASAVSWVSWLPKLKGI
IRKTYVCSHPPTVISEFASPDSESYRIPLNSRNROYSVWVAATISAGRNSGII
TRVELAKAPARILTFSGVTTPMKMDIVLCKAVGDPSPAKMKDMSQPTSLVTDG
RSIFNSGSEFIRTVKAEDSGYSCIANNNMSDEIILNOVPPDOPRLYTSKTS
SSITLSMLPGDNGSGSIRGYILOYESNSEMGSPFIPSESRYLENLCGMYKFT
LTANONGCPRISEITIAKTLCKEPOPSKDELPASINTIRYANLIGNNDGCPRTS
FLEIRPFGITVTTAORTSLSKSYLIDYQEAFTWELQVNCNSAGCEKQANFTL
NYGSEITPLIKSVONNEGLTNEGKMLVTISCLVGLLLEFVLLVYVRRROR
LKRLDKASLAEMLMSKNTRTDLSKQOQTLRMHIDIPRAQLLIERPMTETIDRS
TVLLTDPEAKOKSLTVTHVYOSVATGPGLVDVSDAGTNPTRRARKG
TARRVYASOMTLNRPPTISAHLLTIDWRLPTPRAAGSVKDESDSYSVSPDSDTAR
SNWSTPSASSTYELERATLEHAKMEQLPNAKFTTTECFISDTSEQLTAGNENITD
SLTSSTPSESGICRFTASPPKPDGGGVKMMNAVPAIGVYSTICLHTLEWTFC"

BASE COUNT

1621 a 1658 c 1548 g 1283 t

ORIGIN

Query Match 100.0% Score 20; DB 11; Length 6110;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ccaattctcaagagcagg 20
|||||
Db 4383 CCAGTTCCTCAAGAGACAGG 4402

RESULT 2
AF023450
LOCUS AF023450 6413 bp mRNA PRI 01-JUN-1998
DEFINITION Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (DSCAM)
mRNA, complete cds.
ACCESSION AF023450
VERSION AF023450.1 GI:3169767
KEYWORDS

SOURCE

ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiinae; Homo.
1 (bases 1 to 6413)

REFERENCE

AUTHORS

TITLE

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6413)
AUTHORS Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,
Lyons, G.E. and Korenberg, J.R.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los
Angeles, CA 90048-1869, USA
FEATURES
SOURCE Location/Qualifiers
1. 6413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21 (trisomy 21)"
/map="21q22, between HMG14 and MX1"
/tissue_type="brain"
/dev_stage="14 weeks, fetal"
/clone="CHD2-52"
/note="derived from alternately-spliced mRNA"
1. 6413
/gene="DSCAM"
453. 5168
/gene="DSCAM"
/note="member of immunoglobulin superfamily; involved in
nervous system development"
/codon_start=1
/product="Down syndrome cell adhesion molecule"
/protein_id="AAC17967.1"
/db_xref="GI:3169768"
/translation="IMLILALSLFQSFANFSEDLHSSLYPVNASLQEVFASTTGTLV
PCPAGGIPVYLRMYLATGEITVPGIRHVPNGTLQIFPPSPSTLIHNTYIC
TAENPAGIRSDOVHAKVALREPTVREDDKTRGNVAFAKCIIPSSVAYIIVSW
KDVSVLSGSRFLITSGALYIKDVONEDGLYNYRCITRHRVYGETROSNASALF
DPANSPALDDFVSDPANSPALDDFVSDPANSPALDDFVSDPANSPALDDFVSD
KDVSVLSGSRFLITSGALYIKDVONEDGLYNYRCITRHRVYGETROSNASALF
CLIENTRPSDSGTYVCEVSNRYGTAVIGRLYKOPILATISPRKXSSGVSVLS
CSVTGEQDELISWRNCELINPKGNVITGIIHNLIMDHVKSDBGAYOCFVRKDKL
SAOYVOYVLEDDGPKLISAFSEKVVSPAEFVSLMCNVKGPILPTITWTDDEPIL
GSHISQMITSEGVAWSYLNSSQVBDGYYRGTANNAGVLYQARINVRPASIR
PKNTITAGADDTYIHCRVIGPYYSIKWYKNSNLPEFNHROVALFENNGLKLSVQ
EVDGECVTCNLTNPOLSTOSVHTVAKVPFIOPFEFEPFISIGORVFIQVYVSGDL
PTITWOKGRIIPGSLGVTIDNIDFTSLRLSNLIMHNGNNTCIRARNAVAEHOSOL
QIVRVPKFEVQPDODGYIKRAVILNCSAEQYVPTIYVRFKSGVGFQFOPIALN
GRIOVLSNGSLIKHVAVEEDSGITLCKVSNVGVADVSKSMILYIKIIPAMITSTP
ATCGCKREMSCTAHGEKPIIYVWEKEDRIINPEMARLVSTKEGEVISTLQILP

exon /evidence-not_experimental
11509..11621 exon prediction, score = 74% (0%)
/note="Xpound exon prediction, score = 74% (0%)"
/evidence-not_experimental
exon /note="GRAIL, score = 58.000%, comment = good shadow"
/evidence-not_experimental
12393..12480
/rpt_family="L2"
repeat_region /evidence-not_experimental
complement(12492..12869)
/rpt_family="L1MB3"
/evidence-not_experimental
13831..14128
/rpt_family="AluSx"
repeat_region /evidence-not_experimental
15597..15646
/rpt_family="L2"
exon /evidence-not_experimental
complement(15672..15819)
/note="WZEF, score = 53.3%"
/evidence-not_experimental
15717..15762
/rpt_family="PABL-A"
repeat_region /evidence-not_experimental
complement(17207..17470)
/rpt_family="AluSx"
/evidence-not_experimental
complement(17433..17534)
/note="GRAIL, score = 87.000%, comment = excellent"
exon /evidence-not_experimental
complement(18789..19089)
/rpt_family="AluY"
repeat_region /evidence-not_experimental
complement(19490..20236)
/rpt_family="L1PA15"
repeat_region /evidence-not_experimental
complement(20244..26377)
/rpt_family="L1PA2"
exon /evidence-not_experimental
complement(20451..22019)
/note="GRAIL, score = 92.000%, comment = excellent"
exon /evidence-not_experimental
complement(21819..21970)
/note="WZEF, score = 65.5%"
exon /evidence-not_experimental
complement(22035..22267)
/note="GRAIL, score = 59.000%, comment = good"
exon /evidence-not_experimental
complement(22537..22859)
/note="GRAIL, score = 49.000%, comment = marginal"
exon /evidence-not_experimental
complement(23004..24272)
/note="GRAIL, score = 87.000%, comment = excellent"
repeat_region /evidence-not_experimental
complement(26405..27896)
/rpt_family="L1PA15"
exon /evidence-not_experimental
complement(26651..26723)
/note="GRAIL, score = 50.000%, comment = good"
exon /evidence-not_experimental
complement(26651..26773)
/note="WZEF, score = 52.1%"
repeat_region /evidence-not_experimental
27887..29472
/rpt_family="L1P"
exon /evidence-not_experimental
28426..28580
/note="WZEF, score = 63.6%"
repeat_region /evidence-not_experimental
29473..29835
/rpt_family="THE1B"
/evidence-not_experimental

repeat_region 29836..29969
/rpt_family="L1P"
/evidence-not_experimental
complement(30121..30486)
repeat_region /rpt_family="THE1C"
/evidence-not_experimental
complement(30940..31299)
/rpt_family="L1MCI"
/evidence-not_experimental

Query Match 100.0%; Score 20; DB 11; Length 145861;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 1 ccaggtctcaaggagcagg 0
Db 56809 CCAGTCTCAAGGAGCAGC 56828
|||||

RESULT 4
AF064865/c AF064865 159424 bp DNA PRI 02-JUN-1998
LOCUS Homo sapiens chromosome 21q22.3 PAC 58D10, complete sequence.
DEFINITION AF064865
VERSION AF064865.1 GI:3171160
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 159424)
Bleichschmidt, K., Bagand, E., Hildmann, T., Nordsiek, G., Drescher, B.,
Weber, J., Schattevoy, R., Yaspo, M.-L. and Rosenthal, A.
Direct Submission
TITLE Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
LOCATION/Qualifiers
1..159424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
complement(1224..1523)
/rpt_family="AluY"
/evidence-not_experimental
1263..1386
/note="GC score = 8.50 (124bp)"
/note="Region: GC content"
/evidence-not_experimental
1932..2047
/note="WZEF, score = 92%"
/evidence-not_experimental
3299..3348
/note="Xpound exon prediction, score = 72% (0%)"
complement(3508..3614)
/note="GenScan, score = -0.41%, comment = Terminal_exon
107 bp frame: 0 phase: 2"
/evidence-not_experimental
3843..4124
/note="GRAIL, score = 53.000%, comment = good shadow"
/evidence-not_experimental
complement(3913..4018)
/note="GenScan, score = 16.83%, comment = Initial_exon 106
bp frame: 1 phase: 1"
/evidence-not_experimental
complement(3913..4089)
/note="Xpound exon prediction, score = 94% (0%)"
complement(3913..4066)
/note="WZEF, score = 72%"
/evidence-not_experimental
complement(3913..4062)

repeat_region
misc_feature
exon
exon
exon
exon
exon

exon /note="GRAIL, score = 93.000%, comment = excellent"
 /evidence-not_experimental
 complement(4787..4829)
 /note="Xpound exon prediction, score = 73% (0%)"
 /evidence-not_experimental
 complement(7468..7853)
 /note="Genscan, score = 35.77%, comment = Terminal_exon
 386 bp frame: 0 phase: 2"
 /evidence-not_experimental
 complement(7529..7853)
 /note="GRAIL, score = 86.000%, comment = excellent"
 /evidence-not_experimental
 complement(7546..7853)
 /note="Xpound exon prediction, score = 98% (0%)"
 /evidence-not_experimental
 complement(7563..7853)
 /note="M2EF, score = 72.2%"
 /evidence-not_experimental
 complement(8472..8544)
 /note="GRAIL, score = 49.000%, comment = marginal"
 /evidence-not_experimental
 complement(8772..8977)
 /note="GRAIL, score = 60.000%, comment = good"
 /evidence-not_experimental
 complement(8780..8977)
 /note="Genscan, score = 1.30%, comment = Internal_exon 198
 bp frame: 2 phase: 0"
 /evidence-not_experimental
 complement(9028..9216)
 /note="M2EF, score = 96.1%"
 /evidence-not_experimental
 complement(9028..9216)
 /note="GRAIL, score = 92.000%, comment = excellent"
 /evidence-not_experimental
 complement(9028..9216)
 /note="Genscan, score = 13.54%, comment = Internal_exon
 189 bp frame: 1 phase: 0"
 /evidence-not_experimental
 complement(9842..9963)
 /rpt_family="MIR"
 /evidence-not_experimental
 11055..11354
 /rpt_family="AluY"
 /evidence-not_experimental
 11606..11795
 /rpt_family="MER63A"
 /evidence-not_experimental
 12023..12324
 /rpt_family="AluSX"
 /evidence-not_experimental
 complement(12784..12882)
 /note="M2EF, score = 96.1%"
 /evidence-not_experimental
 complement(12784..12882)
 /note="Genscan, score = 0.56%, comment = Internal_exon 99
 bp frame: 1 phase: 0"
 /evidence-not_experimental
 12814..12872
 /note="GRAIL, score = 59.000%, comment = good shadow"
 /evidence-not_experimental
 complement(13508..13527)
 /note="Xpound exon prediction, score = 67% (0%)"
 /evidence-not_experimental
 complement(14478..14641)
 /note="Genscan, score = 9.80%, comment = Internal_exon 164
 bp frame: 0 phase: 2"
 /evidence-not_experimental
 complement(14478..14590)
 /note="Xpound exon prediction, score = 86% (0%)"
 /evidence-not_experimental
 complement(14478..14641)
 /note="GRAIL, score = 94.000%, comment = excellent"
 /evidence-not_experimental

exon complement(14478..14645)
 /note="M2EF, score = 86.2%"
 /evidence-not_experimental
 complement(16046..16163)
 /note="Genscan, score = 6.30%, comment = Internal_exon 118
 bp frame: 0 phase: 1"
 /evidence-not_experimental
 16257..16330
 /note="GRAIL, score = 43.000%, comment = marginal shadow"
 /evidence-not_experimental
 18126..18259
 /rpt_family="MER5A"
 /evidence-not_experimental
 19077..19876
 /rpt_family="L2"
 /evidence-not_experimental
 19945..20070
 /rpt_family="L1MC3"
 /evidence-not_experimental
 20072..20405
 /note="M2EF, score = 84.2%"
 /evidence-not_experimental
 20106..20254
 /rpt_family="L2"
 /evidence-not_experimental
 complement(20145..20172)
 /note="Xpound exon prediction, score = 73% (0%)"
 /evidence-not_experimental
 complement(20491..20855)
 /rpt_family="THE1B"
 /evidence-not_experimental
 20859..20899
 /rpt_family="L2"
 /evidence-not_experimental
 21282..21348
 /rpt_family="L2"
 /evidence-not_experimental
 21987..22074
 /note="M2EF, score = 85.6%"
 /evidence-not_experimental
 complement(22597..22750)
 /note="M2EF, score = 99.6%"
 /evidence-not_experimental
 complement(22597..22750)
 /note="Xpound exon prediction, score = 87% (0%)"
 /evidence-not_experimental
 complement(22597..22750)
 /note="Genscan, score = 13.62%, comment = Internal_exon
 154 bp frame: 1 phase: 1"
 /evidence-not_experimental
 complement(22597..22750)
 /note="GRAIL, score = 100.000%, comment = excellent"
 /evidence-not_experimental
 22995..23787
 /rpt_family="L1M4"
 /evidence-not_experimental
 23793..23859
 /rpt_family="L1MB8"
 /evidence-not_experimental
 complement(23861..24225)
 /rpt_family="THE1B"
 /evidence-not_experimental
 24241..24582
 /rpt_family="L1MB8"
 /evidence-not_experimental
 24395..24867
 /rpt_family="L1M4"
 /evidence-not_experimental
 24869..25346
 /rpt_family="MER4A2"
 /evidence-not_experimental
 25359..26535
 /rpt_family="L1M4"

exon
/evidence=not experimental
26464..26634

Query Match 100.0%; Score 20; DB 11; Length 159424;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccagttctcaaggagcagg 20
|||||

Db 7849 CCAGTTCTCAAGGACGAGG 7830

RESULT 5
CHKP105A 3654 bp mRNA VRT 28-Apr-1993
DEFINITION Chicken p105 (NF-kappaB p50 precursor) mRNA, complete cds.
ACCESSION M86930
VERSION M86930.1 GI:212525
KEYWORDS NF-kappaB p50 subunit precursor; p105 gene.
SOURCE Gallus gallus (library: cDNA lambda Zap) embryo fibroblasts cDNA to mRNA.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 3654)
Capobianco,A.J., Chang,D.D., Mostafaei,G. and Gilmore,T.D.
p105, the NF-kappa-B p50 precursor protein, is one of the cellular
proteins complex with the v-rel oncoprotein in transformed chicken
splenic cells
J. Virol. 66, 3758-3767 (1992)
FEATURES
MEDLINE 92260550
SOURCE Location/Qualifiers
1..3654
/organism="Gallus gallus"
/db_xref="taxon:9031"
/cell_line="CEF, primary fibroblasts"
/dev_stage="embryo"
/tissue_type="fibroblasts"
/issue_1ib="cDNA lambda zap"
1..217
218..3172
/gene="p105"
218..3172
/note="NF-kappaB p50 subunit precursor"
/codon_start=1
/product="p105"
/protein_id="AA49000.1"
/db_xref="GI:212526"
/translation="MAGEDPYIMGVSDPQMFAMQDLGMSTIFNNNGYITSDLPRLTA
DGPYLOIIEOPKRGFRFRVYCEGSPHGLPGASSEKKNKSPQVAKICNVYGPAAKIV
OLITNGKRYRLHSLVGLKPCEDGVCYNAGPKDMVYGFANLGLIHTKKVFEETLT
RMIDACKRGYNPGLIHPGLGLOAEGCGDROLREEREIIRAOAAYOOTEMDSVVR
LMFTAFIDPSNGGFRRLDPIYSDAIYASKAPNANLKIYMDRTACGVGGEIYLL
CDKVOVDIQRITFEDENGMMWEGFDFSTVHRQPAIYFKTIPKIRDVNITRPAY
FVQLARRKSDLETSEPKPLIYPEIKDKKEVQRKKQKLPNPSDYGSGSGAGGGMG
GGGAGSGGSESYSGYSAFGMHFPGTTSNAGMKHELNSVKKDEESSDQSK
WTKHKDVEETEKNECRTSGHNEKEDASLCCDEGKPKCGODGLFEKAMQLAK
HCNRLPDYAVTGDVRLAVORHLTAVO..ONGNV..CHL:1IHLRELVKMLLEVMD
KINYNNIIMNRDLYOTPLHLAVITRKAQVVDLLKAGANNVLLDRHGNVYLHAAAG
DOKLITLLKHOKASMTIDLSNGELSAIHWVANSISCLLIIAGDVNNOEOKS
GRTALHLAVEQENPLAGCLLLEGDAVDSTTYGTPPLHIAARGFTKRLAAVKAAG
ADPHVNEPLFEDVEEDKDDDDGIVGTPPLDMAMNEVYDILNGKPIIAAAS
EDLISGCPRLNESSKOOLYKLETPPSKMWSTLAEKILGILNNAFLQSPSPST
LIDNKKISGGTGOELIAAFTQMDHTEALEVIQKALLSSROSHOEDKTIKAFPLSPST
SFAKETGEIYNHKKFQPESTCDSCVESFRKLSFTTSDLSNKSITLSMNLTYOR
KAAQKAVYILTR"

BASE COUNT 1097 a 793 c 922 g 842 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 4; Length 3654;
Best Local Similarity 94.7%; Pred. No. 38;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ccagttctcaaggagcagg 20
|||||

Db 2418 CAGTTCTCAAGGACGAGG 2436

RESULT 6
GGAF000241
LOCUS GGAF000241 3846 bp mRNA VRT 28-MAY-1997
DEFINITION Gallus gallus nuclear factor NF-kB1 mRNA, complete cds.
ACCESSION AF000241
VERSION AF000241.1 GI:2130627
KEYWORDS chicken.
SOURCE Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 3846)
Cabannes,E., Vives,M.-F. and Bedard,P.-A.
Transcriptional and post-transcriptional regulation of
kB-controlled genes by pp60v-src
Oncogene (1997) in press
2 (bases 1 to 3846)
Bedard,P.-A., Cabannes,E. and Vives,M.-F.
Direct Submision
Submitted (18-Apr-1997) Biology, York University, 4700 Keele
Street, North York, Ontario M3J 1P3, Canada
FEATURES
SOURCE Location/Qualifiers
1..3846
/organism="Gallus gallus"
/db_xref="taxon:9031"
5'UTR 1..160
CDS 161..3178
/codon_start=1
/product="nuclear factor NF-kB1"
/protein_id="AA58343.1"
/db_xref="GI:2130628"
/translation="MEATATATBARLNPQCADPGILKMGEDPYINGVSDPQMFAMQDL
GMSTIFNNNGYITSDLPRLTAADGPYLOIIEOPKRGFRFRVYCEGSPHGLPGASSEK
NKSYPOVKICNVGPAKIVYQLVNGLYVHLSHSLVGLKPCEDGVCYNAGPKDMVY
GFANGLIHTKKVFEETLTPEIMIDACKRGYNPGLIHPGLGLOAEGCGDROLRE
RETIKQAAVOOTKENDLSVRLMFTAFIDPSNGGFTTRRLDPIYSDAIYASKAPNANL
KIVRMORTAGCVTGEIEIYLLCDKVOKDIOIRFEDENGMMWEGFDFSTVHRQK
FAIVKTPKRYRDVNITKPAVSFVQLARRKSDLETSEPKPLIYPEIKDKKEVQRKKQK
LPNPSDYGSGSGAGGGMGCGGAGSGSESYSGYSAFGMHFPGTTSNAGMKHELNSV
KDEESSDQSKWTKHKDVEETEKNECRTSGHNEKEDASLCCDEGKPKCGODGLFEKAM
QLAKHCNRLPDYAVTGDVRLAVORHLTAVO..ONGNV..CHL:1IHLRELVKMLLEVMD
KINYNNIIMNRDLYOTPLHLAVITRKAQVVDLLKAGANNVLLDRHGNVYLHAAAG
DOKLITLLKHOKASMTIDLSNGELSAIHWVANSISCLLIIAGDVNNOEOKS
GRTALHLAVEQENPLAGCLLLEGDAVDSTTYGTPPLHIAARGFTKRLAAVKAAG
ADPHVNEPLFEDVEEDKDDDDGIVGTPPLDMAMNEVYDILNGKPIIAAAS
EDLISGCPRLNESSKOOLYKLETPPSKMWSTLAEKILGILNNAFLQSPSPST
LIDNKKISGGTGOELIAAFTQMDHTEALEVIQKALLSSROSHOEDKTIKAFPLSPST
SFAKETGEIYNHKKFQPESTCDSCVESFRKLSFTTSDLSNKSITLSMNLTYOR
KAAQKAVYILTR"

BASE COUNT 1167 a 815 c 939 g 925 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 4; Length 3846;
Best Local Similarity 94.7%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ccagttctcaaggagcagg 20
|||||

Db 2424 CAGTTCTCAAGGACGAGG 2442

RESULT 7
CHKNFKB1 4023 bp mRNA VRT 03-FEB-1999

DEFINITION Chicken mRNA for NF-kB p50 subunit.
ACCESSION D13719
VERSION D13719.1 GI:222838
KEYWORDS DNA binding protein; NF-kB p50; NF-kB/rel family.
SOURCE Gallus gallus (library: lambda gtl1) spleen, cDNA to mRNA, clone pCR410.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 4023)
Ikeda,T.
REFERENCE Direct Submission
AUTHORS Submitted (19-NOV-1992) to the DDBJ/EMBL/GenBank databases. Toshio
JOURNAL Ikeda, Institute for Virus Research, Kyoto University; Shogoin,
TITLE Sakyo-Ku, Kyoto 606, Japan (E-mail: tikedae@virus.kyoto-u.ac.jp,
tel:075-751-3990, Fax:075-751-3991)
2 (bases 1 to 4023)
Ikeda,T., Honjo,K., Hirota,Y. and Onodera,T.
REFERENCE Isolation of the chicken NF-kappa B p65 subunit-encoding cDNA and
AUTHORS characterization of its products
JOURNAL Gene 133 (2), 237-242 (1993)
COMMENT Submitted (19-NOV-1992) to DDBJ by:
Toshio Ikeda
Institute for Virus Research
Kyoto University
Shogoin, Sakyo-ku
Kyoto 606
Japan
Phone: 075-751-3990
Fax: 075-751-3991
E-mail: tikedae@virus.kyoto-u.ac.jp.
Location/Qualifiers
1..4023
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone_lib="lambda gtl1"
/issue_type="spleen"
404..3355
/codon_start=1
/product="NF-kB p50 subunit"
/protein_id="BAA02872.1"
/db_xref="GI:222838"
/translation="MAGEDPPIKGVSDPQMFAMDQLMNSTIFNNITGYITSDPLRTA
DGPYIOTIEPKQKGRFRFRVCEGSPHGGIPGASSSEKNSKYPQVVICNVYPAYIV
QLVTKNGKYVHLHSLVGRFCDEGCVTVNAGPKDVMVGANGLIHLTKKYELET
RMIDACKKGYNPGFLVHPELGYDQAGCGDROLTEREERTIKRAVQQRKEDLSVVR
LMFTAFLPDSNGGFTRLDPVLSDAIYDSKAPNASLKIIVRMDRTAGCVTGEETLL
CDVVOVDIOIRFYEDENGAMGEGFDESPDVHROFAIVETPKYRDVNTTKPASF
FQLRKRSQDTESEKPPLYPEIKDKEVQRKRLMNEFSGYGGSGAGGCMFG
GGGGGAGSGTSYPSYGSYSAFGCMHFRGTTKSNAGKHLNSITVKKDESSKOSDK
WDTRHDVKVETVEKNEORTSGHNEKEEDASLCKDGNKPKCCGCGCLFEKAMOLAK
RHONALFDYAVTGDVRLAVORHLAVODNGDNVHLSTIHLHELKYNLLEVPAD
MANYNIIIMNRNDLYOTPLHLAVITKQAEVEDLLKAGANNLDRHNSVHLAAEG
DDKILSLLEKHKASSMIDLSNGEGSAIHVVYANSLSCLKLLIAGVDVANAQEQS
GRTAHLAVEOENIPLAGCLLEGDVADSTYDGTTPHLIAGRGFTKLAIVLKAG
ADPHVENPEFLPVEEDVDKDDDEDPGTPPLDAMANEVVDIINAKPPIYAAVS
EDLISGSPLELNESSKOQYKXILFPPDSKMSSTLAERKLGILNNAPRISPSPT
LIDNRTISGTVDELIALTOMDHTETAEIYOKALSSSORQIYOEDNTIAPPSISPT
SFAKEETGELYNHFKFDPPESTCDSGVETSFRLKFTYSDLSNKSSTITLSKMTLGICQ
ESSYOSSYIPN"
BASE COUNT 1188 a 866 c 997 g 972 t
ORIGIN
polya-signal 4011..4016
Query Match 87.0%; Score 17.4; DB 4; Length 4023;
Best Local Similarity 94.7%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 8
AL133354/c
LOCUS
DEFINITION
ACCESSION AL133354
VERSION AL133354
KEYWORDS
SOURCE HTGS_PHASE1.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 113575)
REFERENCE Direct Submission
AUTHORS Submitted (03-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00010 Length: 2739bp
Contig_ID: 00025 Length: 2967bp
Contig_ID: 00028 Length: 2249bp
Contig_ID: 00035 Length: 1470bp
Contig_ID: 00036 Length: 1428bp
Contig_ID: 00039 Length: 1821bp
Contig_ID: 00055 Length: 1174bp
Contig_ID: 00064 Length: 1448bp
Contig_ID: 00065 Length: 5014bp
Contig_ID: 00083 Length: 1029bp
Contig_ID: 00096 Length: 1723bp
Contig_ID: 00107 Length: 1027bp
Contig_ID: 00124 Length: 1596bp
Contig_ID: 00126 Length: 3878bp
Contig_ID: 00141 Length: 1080bp
Contig_ID: 00150 Length: 1820bp
Contig_ID: 00193 Length: 1075bp
Contig_ID: 00209 Length: 3504bp
Contig_ID: 00224 Length: 2681bp
Contig_ID: 00250 Length: 1771bp
Contig_ID: 00253 Length: 1431bp
Contig_ID: 00278 Length: 1129bp
Contig_ID: 00292 Length: 1380bp
Contig_ID: 00327 Length: 1536bp
Contig_ID: 00329 Length: 1516bp
Contig_ID: 00344 Length: 1814bp
Contig_ID: 00354 Length: 4921bp
Contig_ID: 00367 Length: 1147bp
Contig_ID: 00409 Length: 1143bp
Contig_ID: 00418 Length: 3388bp
Contig_ID: 00429 Length: 3151bp
Contig_ID: 00434 Length: 1915bp
Contig_ID: 00437 Length: 1068bp
Contig_ID: 00480 Length: 2277bp
Contig_ID: 00491 Length: 1303bp
Contig_ID: 00500 Length: 1398bp
Contig_ID: 00542 Length: 2072bp
Contig_ID: 00564 Length: 2215bp
Contig_ID: 00570 Length: 1152bp
Contig_ID: 00571 Length: 1867bp
Contig_ID: 00579 Length: 2258bp
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source 1..113575
Location/Qualifiers

Thu Apr 27 14:22:29 2000

us-08-956-991-2.rst

Page 1

OM of: US-08-956-991-2 to: EST:* out-format: pfs

Date: Apr 25, 2000 9:53 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=farme-p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US08956991/runat_25042000_121637_9550/app-query.fasta.1
-DB=EST -PMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MIMATCH=0.100 -LODPC=10.000 -LODPEXT=0.000 -EGAPOP=6.000
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR_SCORE=pcpt -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NOR=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08956991
-NCPU=6 -ICPU=3 -NO_XLPRX -WAIT -THREADS=1

Search information block:

Query: US-08-956-991-2

Query length: 1910

Database sequences: 4538634

Database length: 1887831982

Search time (sec): 2119.170000

score_list:

Sequence	Strid	Orig	ZScore	EScore	Len
gb_est27:AI454704	-	723.00	1392.53	8.5e-69	458
gb_est27:AI454704	-	665.00	1279.55	1.7e-62	427
gb_est18:AJ0032472	-	504.00	1162.31	5.6e-56	355
gb_gss1:FR0022941	-	576.00	1099.35	1.8e-52	619
gb_est27:AI454704	-	446.00	853.74	8.7e-39	310
gb_est27:AI454704	-	420.00	823.05	4.4e-37	294
gb_est27:AI454704	-	410.00	797.51	1.2e-35	510
gb_est27:AI454704	-	411.00	774.49	2.3e-34	466
gb_est18:AM048129	-	408.50	776.34	1.8e-34	466
gb_est18:AM048129	-	408.50	766.78	6.1e-34	466
gb_gss3:BO7180	-	379.50	717.27	3.5e-31	455
gb_est27:AI454704	-	355.50	667.23	2.1e-28	569
gb_gss1:FR0021997	-	346.50	648.55	3.8e-27	610
gb_est28:AI542894	-	345.50	644.70	3.8e-27	606
gb_est17:AM014717	-	344.00	631.34	2.1e-27	498
gb_est17:AM014717	-	332.50	619.73	2.1e-26	298
gb_gss1:FR0030733	-	329.00	616.95	1.3e-25	442
gb_gss1:FR0030733	-	317.50	590.98	3.7e-24	635
gb_gss1:FR0021994	-	298.50	554.34	4.1e-22	610
gb_gss1:FR0030733	-	293.50	544.30	1.5e-21	614
gb_est28:AI498283	-	289.50	540.87	1.5e-21	585
gb_est27:AI406159	-	285.00	528.28	1.2e-20	585
gb_gss1:FR0030738	-	283.50	524.86	1.3e-20	605
gb_gss1:FR0030738	-	275.50	509.21	1.3e-19	602
gb_gss1:FR0030738	-	274.00	502.47	3.2e-19	785
gb_gss1:FR0030738	-	272.00	499.67	3.1e-19	597
gb_est12:AI107879	-	250.00	459.67	1.7e-17	578
gb_est12:AI107879	-	246.50	452.01	1.1e-16	570
gb_gss1:FR0022007	-	246.00	452.01	1.1e-16	570
gb_est28:AI543452	-	240.00	450.26	2.6e-16	282
gb_est14:CN506594	-	231.50	414.13	2.7e-14	1101
gb_gss1:CN506594	-	226.50	408.57	5.4e-14	816
gb_est20:AA817246	-	224.00	400.80	1.5e-13	568
gb_gss1:FR0022012	-	224.00	400.80	1.5e-13	568
gb_gss1:CN501040	-	221.00	402.21	9.8e-14	947
gb_est12:AI131274	-	221.00	402.21	9.8e-14	947
gb_gss1:HSK18812	-	221.00	401.34	2.3e-14	147
gb_gss1:FR0009351	-	220.00	401.34	2.3e-14	238
gb_est14:AM213058	-	215.00	383.59	1.3e-12	532
gb_gss1:CN5012193	-	215.00	383.59	1.3e-12	1033
gb_est28:AI513189	-	215.00	387.94	7.6e-13	711
gb_est26:AU004083	-	215.00	387.94	7.6e-13	711

gb_gss1:FR0022080 + 214.50 390.77 5.3e-13 544 ! AL014951 F.rubripes GSS sequ
gb_est12:AI722009 + 214.00 388.51 7.1e-13 595 ! AI722009 fd18c09.y1 Sugano K
gb_gss1:CN50070A + 208.00 368.92 8.8e-12 1028 ! AL067489 Drosophila melanog

seq_name: gb_est27:AI454704

seq_documentation_block: 458 bp mRNA EST 05-JUL-1999
LOCUS AI454704
DEFINITION UI-R-BT0-qk-h-03-0-UI.s1 UI-R-BT0 Rattus norvegicus CDNA clone
VERSION AI454704
KEYWORDS AI454704.1 GI:4295587
SOURCE EST
ORGANISM Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 458)
REFERENCE
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
On Mar 20, 1998 this sequence version replaced gi:2980496.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iue.wiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics (www.resgen.com) This clone is also
available through the I.M.A.G.E. Consortium at LNL
(info@image.llnl.gov) IMAGE ID=1788947
Seq primer: M13 forward
POLYA-No.

FEATURES
source
Location/Qualifiers
1..458
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT0-qk-h-03-0-UI"
/clone_lib="UI-R-BT0"
/dev_stage="adult"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; This library
(UI-R-BT0) consists of a mixture of individually tagged
normalized libraries constructed from rat hippocampus,
thalamus, mid-brain, medulla, corpus striatum, cerebral
cortex and testis. The tag used to identify the source
tissue is a string of 3-6 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture. This library was then subcloned using a
driver consisting of a mixture of all clones from UI-R-A0,
UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and
UI-R-C2P."

BASE COUNT 107 a 112 c 123 g 116 t
ORIGIN
alignment_scores: 723.00 Length: 152
Quality: 5.021 Gaps: 0
Percent Similarity: 94.737 Percent Identity: 91.447

alignment_block:
US-08-956-991-2 x AI454704/rev

Align seg 1/1 to reverse of: A1454704 from: 1 to: 458

```
525 CysArgValIleGlyTyrProTyrTyrSerIleIleSTyPylLysAsnE 541
|||||
457 TCCTGTGATGTTGTTACCTTATTACTCATCAAGTGTACAGAACGA 408
541 rAnLeuLeuProPheAsnHisArgIleValAlaPheGluAsnAsnGlyT 558
|||||
407 TAACCTGCTTCTTTCAACACCGACGAGGTGGCGTTTGAGAACATGTGA 358
558 hrLeuLysLeuSerAspValGlnLysGluValAspGlnGlyLutTyrThr 574
|||||
357 CTCTGAACCTTCAGATGTGAGAAAGAGTGTGACGAGGAGAGTACACA 308
575 CysAsnValLeuValGlnProGlnLeuSerThrSerGlnSerValHisVa 591
|||||
307 TGTACCTGTGTGTGCAGCCACAGCTCTCCACGACGACAGTGTCCACGT 258
591 LThrValLysValProPheIleGlnProPheGluPheProArgPheS 608
|||||
257 GACAGTCAAAGTTCGTCCTTTTATCCACCTTTGAGTTCCGAAATTTT 208
608 erIleGlyGlnArgValPheIleProCysValValValSerGlyAspLeu 624
|||||
207 ATATCGGTGACGGGTATTCATCCCATGTGTGTGTCTCGGGGACTTA 158
625 ProIleThrIleThrProGlnLysAspGlyArgProIleProGlySerIe 641
|||||
157 CCCATACCATCAGTCGTCGAGAGAGATGCCGTCGATCCCAAGACCTT 108
641 uGlyValThrIleAspAsnIleAspPheThrSerSerLeuArgIleSerA 658
|||||
107 CGGTGTACCATGACACACATGACATTCACACGCTCTTGAGAGATTTCA 58
658 snLeuSerLeuMetHisAsnGlyAsnTyrThrCysIleAlaIArgAsnGlu 674
|||||
57 ACCTCTCAGTATGATGACAAAGGGAATACACATGATCGAGCGGAAG 8
675 AlaAla 676
|||||
7 GCGGCC 2
```

seq_name: qb_est41:AM159035

seq_documentation_block: 427 bp mRNA

LOCUS AM159035 2830605.x1 Xenopus EST library Xenopus laevis cDNA clone za50605

DEFINITION 5', mRNA sequence.

ACCESSION AM159035

VERSION AM159035.1 GI:6271064

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;

Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae;

Xenopus.

REFERENCE 1 (bases 1 to 427)

AUTHORS Schütz, K., de la Bastide, M., Huang, E.-N., Nascimento, L., Preston, R.,

Shah, R., Swaby, I., Shekher, M., Spiegel, L., Vil, M.D. and

McCombie, W.R.

TITLE Expressed sequence tags from Xenopus

JOURNAL Unpublished (1999)

COMMENT On May 18, 1998 this sequence version replaced gi:3137856.

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: za50 row: e column: 05

Seq primer: M13 universal forward primer

High quality sequence stop: 427.

FEATURES

Source

Location/Qualifiers

1..427

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="za50605"

/clone_lib="Xenopus EST library"

/tissue="total brain tissue"

/cell_line="W22-TGA"

/dev_stage="taopole"

/note="Vector: Lambda zap I; Site-1: XbaI; This library

was supplied by Holly Cline (Cold Spring Harbor Labs).

cDNA synthesis with oligo dt xba I (xba I cloning site).

RNA: stage 50-56 tadpoles, total brain tissue, GTC

extraction method."

BASE COUNT

114 a 122 c 91 g 100 t

ORIGIN

alignment_scores:

Quality: 665.00

Ratio: 4.890

Percent Similarity: 97.842

Percent Identity: 92.806

alignment_block:

US-08-956-991-2 x AM159035

Align seg 1/1 to: AM159035 from: 1 to: 427

410 lIleSerAlaPheSerGluLysValSerProAlaGluProValSerIle 426

|||||

17 ATCTCTCATTTAGCGAGAAAGTGTTAGCCCGGAAACCTGTCTCCCT 66

426 uMetCysAsnValLysGlyThrProLeuProThrIleThrTyrThLeuA 443

|||||

67 TATGTGACAGCTTAAGAGACGCCCTCTCCGACGATCATTGACACCTGG 116

443 sPaSPaSPProIleLeuLysGlyLysSerHisArgIleSerGlnMetIle 459

|||||

117 ACGAGACCCCATATATCAGACGGCCACACAGAAATTAAGTCAAGTAAAC 166

460 ThrSerGluGlyAsnValValSerTyrLeuAsnIleSerSerSerGlnVa 476

|||||

167 ACCTCTGAGGCGACAGTGTGACATTACCTGACATCACCACACTCAAGT 216

476 lArgSPglYglYValTyrArgCysThrAlaAsnAsnSerAlaGlyValY 493

|||||

217 CCGGATATGCGGGGTGTACCGATGACGCGCTACCAACTGTGCGGGCTCG 266

493 aLeuTyrGlnAlaArgIleAsnValArgGlyProAlaSerIleArgPro 509

|||||

267 TCTGTACAGAGCTCGAATAAGCT CCTGCAAGTATCCGTCCA 309

510 MetLysAsnIleThrAlaIleAlaGlyArgAspThrTyrIleHisCysAr 526

|||||

310 ATGAAACATCAGCGGCATAGCGCGCGGACACACATCTCCACTGCGCG 359

526 gValIleGlyTyrProTyrTyrSerIleIleSTyPylLysAsnSerAsn 543

|||||

360 CGTATTCGATATCCCTACTACTCATCAAGTGTGTACAAACATCTAAC 409

543 euLeuProPheAsnHis 548

|||||

410 TCCTCCCATTTAATCAG 426

seq_name: qb_est18:AJ003472

seq_documentation_block:

LOCUS AJ003472 355 bp mRNA

DEFINITION AJ003472 Selected chromosome 21 cDNA library Homo sapiens cDNA

ACCESSION AJ003472

VERSION AJ0 3472.1 GI:2769503

KEYWORDS EST

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS 1 (bases 1 to 355)
Lehrach, H., and Yaspo, M.L.H.
TITLE An integrated transcript map for the whole human chromosome 21
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1288678.
Contact: Yaspo, M.-L.
Max Planck Institut fuer Molekulare Genetik
Inmestrasse 73, D14195 Berlin-Dahlem, Germany
Chr21.subseqec_p112:
FEATURES
source location/Qualifiers
1.355
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q"
/clone="MP112-218"
/clone_lib="Selected chromosome 21 cDNA library"
BASE COUNT 83 a 83 c 90 g 99 t
ORIGIN

alignment_scores:
Quality: 604.00 Length: 115
Ratio: 5.345 Gaps: 0
Percent Similarity: 98.261 Percent Identity: 98.261

alignment_block:
US-08-956-991-2 x AJ003472/rev ..

Align seg 1/1 to reverse of: AJ003472 from: 1 to: 355

1345 TleatgtrrrvalysAlaGlaspsersgltyrtyrserscyllealeaas 1362
|||||
353 ATTCCACAGCGTGAAGACGAGACCTCCGCTATTACAGCTCATTTGCCAA 304
|||||
1362 nasnasntrpqlysersaspgluilleleuasnleuglnvalglnalp 1379
|||||
303 TAACAACTGGGGATCTGATGAATATTATTAACTTACAGTACAGATTC 254
|||||
1379 rorproaspglnproargleuthrvalserlysthrthrserserle 1395
|||||
253 CACCAGATCAGCTCGCTATAGCTCCAGACCCACTCTTCCTCCATC 204
|||||
1395 ThrleusertpleuProglYAspasnglylserseleleatrglyty 1412
|||||
203 ACCCTTTTGGCTCCTCGAGACACAGGGGGCACTTATCAGAGATA 154
|||||
1412 rllleuglntrysersgluaspsnsersgluglntrpqlyserspheproi 1429
|||||
153 CATACTGAGTACTCGAGACAATAGTACAGAGTGGGAGATTTCCTCA 104
|||||
1429 lesersprosergluarserlyrargleuglnuasnleulyscylgtyhr 1445
|||||
103 TCAGCCCGACGGAACGTTCTATGCGTCAATAATCTCAATGTGGGCT 54
|||||
1446 TrpTyrlyspheThrleuThrAlaGlasnglyValGlyProgly 1460
|||||
53 TGGTATAGTTCACACTACAGCCCAAAATGAGTGGGCCAGGG 9
|||||

seq_name: qb_gssl:FR0002941

seq_documentation_block:
LOCUS FR0002941 619 bp DNA GSS 27-FEB-1997
DEFINITION F.rubripes GSS sequence, clone 015G16a10, genomic survey sequence.
ACCESSION Z86724
VERSION Z86724.1 GI:1883636
KEYWORDS GSS: genome survey sequence.
SOURCE Fugu rubripes.
ORGANISM Fugu rubripes

Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.
AUTHORS 1 (bases 1 to 619)
Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranaia, Y.,
Williams, G., and Brenner, S.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgm.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
FEATURES
source location/Qualifiers
1.619
/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 015G16"
/clone="015G16a10"
BASE COUNT 133 a 169 c 141 g 156 t 20 others
ORIGIN

alignment_scores:
Quality: 576.00 Length: 181
Ratio: 4.204 Gaps: 3
Percent Similarity: 75.691 Percent Identity: 66.298

alignment_block:
US-08-956-991-2 x FR0002941 ..

Align seg 1/1 to: FR0002941 from: 1 to: 619

22 SerSerleuTyrPheValAsnAlaSerLeuGlnGluValAlaPheAla 38
|||||
3 TCACGCTTATTATTGTCAATGCATCTGTGCAAGAGAGTGTTCCTCAT 52
|||||
38 TrpThrThr GlyThrleuValProCysProAlaAlaGlyLeuProProVal 54
|||||
53 CACCACAGGAGTACGTTGTCCTGCTGCGGCTGACAGTGTGCCCCCTTC 102
|||||
55 ThrleuArgTrpTyrleuAlaThrGlyGluGluIleTyrAspValProGl 71
|||||
103 TCATPACGGGTGACCTGCTACTGAGAGAGATCTTACGATGTCGAC 152
|||||
71 YlleaighisvalhisproasnclYThrleuglnllephepropepro 88
|||||
153 NATCCGACGTCGACCTCAATGGCACCCCTTCAGATCTTCACATCCAC 202
|||||
88 rosersepheserThrleuIleHisaspasnThrTyrTyrCysThrAla 104
|||||
203 CGTCATATGTCANCAACATCATCATGACACACACTACTACTGCACAGC 252
|||||
105 GluAsnProserGlyLyslleargsergluaspsvalHislelys.... 119
|||||
253 GAGAACCCCTCAGGGAGATCANAGCCNACATGCTCATTAAGGCTGG 302
|||||
119 119

120 TGGCGTCGCGGNNTCACCTGCNNCTCCANATNATTTCCAGACTGACTG 352
|||||
120Ala..ValleuArgLuprot 126
|||||
353 TTCCGAGNCTNNCTGTGTCTCTCCCATNNCCAGTCTCTACAGAGCCCT 402
|||||
126 YrThrValArgValGluAspGlnLysThrMetArgGlyAsnValAlaVal 142
|||||
403 ACACGGTCCGGGTAGAGATCAAAAAGCTATGCGAGGACGCTGCACATC 452
|||||
143 PheLysCyslleleproserSerValGluAlaTyrIleThrValAlaIse 159
|||||

```

453 TTCAAGTCGATTATCCCTGCTAGTGAGNACATATATCATCTGTGTGTC 502
159 rTTPGtLysAspThrValSerLeuValSerGlySerArg 172
||||| |||||||:|||||: |||: |||:
503 CTGGGANAAGACACCATGTCTCAATTAATCTGAAAGTAG 542

seq_name: qb_est2.F13426

seq_documentation_block:
LOCUS F13426 310 bp mRNA EST 15-MAR-1995
DEFINITION HSC2XA021 normalized infant brain cDNA Homo sapiens cDNA clone
c-2xa02, mRNA sequence.
ACCESSION F13426
VERSION F13426.1 GI:710043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 310)
Auffray,C., Behar,G., Bols,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,P., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Poulot,Y.,
Sebastiani-Kabakchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

JOURNAL
MEDLINE
COMMENT Contact: Genethon
GenexPress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
GenexPress_library_idt: C; GenexPress_sequence_idt: y2c-2xa02
Insert Length: 1500 Std Error: 0.00
Seq primer: (-21)M13-universal
High quality sequence stop: 335.
Location/Qualifiers
1. 310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-2xa02"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue-type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: latmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue-type=total
brain; total mRNA was oligo-(dt) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
latmid BA vector. Clone library from B.Soaers, Psychiatry
Dept. Columbia University, USA. Normalization method:
Bento Soares, P.N.A.S in press"
Bento Soares, P.N.A.S in press"

BASE COUNT 79 a 73 c 89 g 66 t 3 others
ORIGIN

alignment_scores:
Quality: 446.00 Length: 103
Ratio: 4.505 Gaps: 0
Percent Similarity: 96.117 Percent Identity: 95.146

alignment_block:
US 08-956-991-2 x F13426 ..
Alkan seq 1/1 to: F13426 from: 1 to: 310
1610 PheValLeuLeuValValArgArgArgArgGluGluArgLeuLy 1626
|||||

```

```

2 TTGTCTCTCTCTGTTGTTCTCCGAGAGCGCGGAGAG. AGCCTAAA 50
1626 sATgLeuAArgAspAlaLysSerLeuAlaGluMetLeuMetSerLyAsnT 1643
||||| |||||||:|||||: |||: |||:
51 GAGCTGCGGAGATGCAAAAGATTACTGAAATNCTCATGAGTAACATA 100
1643 hArArgThrSerAspThrLeuSerLyGluGluGlnGlnThrLeuArgMetHis 1659
||||| |||||||:|||||: |||: |||:
101 CCCGACTTCAGTACGTTAAGCAAGCAGACAGCCTCGCAATGCAC 150
1660 lLeAspIleProArgAlaGlnLeuLeuIleGluGluArgAspThrMetG1 1676
||| |||||||:|||||: |||: |||:
151 ATCGCATACCAAGGCGCTCAGCTTTGATTGAAGAGAGACACCATGGA 200
1676 uThrIleAspAspArgSerThrValLeuLeuThrAspAlaAspPheGlyG 1693
||||| |||||||:|||||: |||: |||:
201 GACCATTGATGATCGCTCCACGGCTCTNTTGACGCGTCTGACTTGGAG 250
1693 LuAlaAlaLysGlnLysSerLeuThrValThrHisThrValHisTyrGln 1709
||||| |||||||:|||||: |||: |||:
251 GGGCAGCTAAGCAGAAAGTCCCTGACGCTCCTCACAGGTCATTACCAA 300
1710 SerValSer 1712
||||| |||
301 TCGGTGTCT 309

seq_name: qb_est1:245894

seq_documentation_block:
LOCUS 245894 294 bp mRNA EST 14-NOV-1994
DEFINITION HSC2MH041 normalized infant brain cDNA Homo sapiens cDNA clone
c-2wh04, mRNA sequence.
ACCESSION 245894
VERSION 245894.1 GI:575128
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 294)
Auffray,C., Behar,G., Bols,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,P., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Poulot,Y.,
Sebastiani-Kabakchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
Contact: Genethon
GenexPress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
GenexPress_library_idt: C; GenexPress_sequence_idt: yle-2wh04
Seq primer: (-21)M13-universal.
Location/Qualifiers
1. 294
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-2wh04"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue-type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: latmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue-type=total
brain; total mRNA was oligo-(dt) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
latmid BA vector. Clone library from B.Soaers, Psychiatry
Dept. Columbia University, USA. Normalization method:
Bento Soares, P.N.A.S in press"
Bento Soares, P.N.A.S in press"

FEATURES
SOURCE

```



```

1164  InvAlbLpAlaPheThrArgAlaGlyAspGlyValArgSerGluGlnIle 1180
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
404  AGGTTTGCCACACCGCTATGGCGATGGAGTGTGTATCAACACCATG 453
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1181  PheThrArgThrLysGluAspValProGlyProPheAlaGlyValLysAl 1197
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
454  TTTGGCCACAGCGAAGCATGTCGCTGAGCTCGCGACGATATCAAGA 503
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1197  aaAlaAlaAlaSerAlaSerMetValPheAlaSerPheIleProIleuL 1214
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
504  TGTATCCAGCTCATCTCAATCACTGATATCTCTGTGGTTCACCAATG 553
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1214  yslLeuAsnGlyIleIleIleArgL} TyrThrValPhe 1225
:::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
554  AACCAGACGGTGTGATCACCAGATACAGTCTTAC 588
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

seq_name: gb_est38:AW048129

LOCUS documentation_block:
      LOCUS      AW048129              402 bp      mRNA      18-SEP-1999
      DEFINITION  UI-M-BH1-ald-g-f-04-0-UI-s1 NIH_BMAP_M.S2 Mus musculus cDNA clone
      VERSION     AW048129
      ACCESSION   AW048129
      KEYWORDS    EST.
      SOURCE      house mouse.
      ORGANISM    Mus musculus.
                  Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Rodentia; Sciurognatha; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      1 (bases 1 to 402)
              Bonaldi,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
COMMENT      On Jun 22, 1998 this sequence version replaced gi:3247317.
              Contact: Chin, H
              National Institute of Mental Health
              6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
              20892-9643, USA
              Tel: 301 443 1706
              Fax: 301 443 9890
              Email: mestr@mail.nih.gov
              The sequence contained an oligo-dT track that was present in the
              oligonucleotide that was used to prime the synthesis of first
              strand cDNA and therefore this may represent a bonafide poly A
              tail. The sequence tag present in the cDNA between the NotI site
              and the oligo-dT track served to identify it as a clone from the
              normalized brain stems library cDNA library preparation: M.B.
              Soares lab Clone distribution: NIH BMAP cDNA clones will be made
              available by the means that is soon to be determined. When NIH
              determines the means for distribution of the BMAP cDNA clones, this
              record will be updated accordingly when that means is determined.
              The following repetitive elements were found in this cDNA sequence:
              1-28, >AT rich low complexity
              Seg primer: M13 Forward
              POLY-A=yes.
FEATURES
      source
          1..402
          /location/qualifiers
          /organism="Mus musculus"
          /strain="C57BL/6J"
          /db_xref="taxon:10090"
          /clone="UI-M-BH1-ald-g-f-04-0-UI"
          /clone_lib="NIH_BMAP_M.S2"
          /dev_stage="27-32 days"
          /lab_host="DHI0B (Life Technologies)"
          /note="Vector: pT733-Pac (Pharmacia) with a modified
          polylinker. Site 1: Not I; Site 2: Eco RI; The
          NIH_BMAP_M.S2 library is a subtracted library derived from
          NIH_BMAP_M.S1, which in turn is a subtracted library
          derived from a mixture of normalized libraries from ten
          regions of the mouse brain (cerebellum, brain stems,

```

olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH.BMAP.M.S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries.

TAG_LIB=NIH_BMAP_M.S2
TAG_TISSUE=brain-stems
TAG_SEQ=TCATG

BASE COUNT 52 a 111 c 129 g 100 t
ORIGIN

alignment_scores:
Quality: 408.50 Length: 131
Ratio: 3.782 Gaps: 1
Percent Similarity: 82.443 Percent Identity: 56.489

alignment_block:

US-08-956-991-2 x AM048129/rev ..

Align seg 1/1 to reverse of: AM048129 from: 1 to: 402

```
383 TyrGlnCysPheValArgLysAspLysLeuSerAlaGlnAspTyrValG1 399
|||||.....:.....:.....:.....:.....:.....:
401 TACGAGTCTTCGCCACGCCGACGCCGACGCCGACGCCGACGCTTGCAT 352
399 nValValLeuGluAspGlyThrProLysIleIleSerAlaPheSerGlu 416
:.....:.....:.....:.....:.....:.....:
351 CATTTGCTGTAAGACGCGACGCCCGCATGCTCTCATCTTCACGCGAGA 302
416 yValValSerProAlaGluProValSerLeuMetCysAsnValLysGly 432
|||||.....:.....:.....:.....:.....:.....:
301 AGGTGCTCAACCTGGGGAGGACGATTCATGATGTGTGGC :CCAAGGCG 252
433 ThrProLeuProThrIleThrThrLeuAspAspAspProIleLeuLy 449
:.....:.....:.....:.....:.....:.....:
251 GCCCGCGCCCGACGGTCACCTGGGCGCTGACGACGACGCGCGTGTGGC 202
449 sGlyGlySerHisArgIleSerGlnMetIleThrSerGluGlnValVal 466
:.....:.....:.....:.....:.....:.....:
201 GGACGAGGACCGCCGACCAACAGTACACATGTCGACGCGACCAACA 152
466 aSerTyrLeuAsnIleSerSerSerGlnValArgAspGlyGlyValTyr 482
:.....:.....:.....:.....:.....:.....:
151 TCAGCCACATGATGACGTCACGGGTCCCGCATCGGAGTGGGGCGGTGAC 102
483 ArgCysThrAlaAsnAsnSerAlaGlyValValLeuTyrGlnAlaArgI 499
|||||.....:.....:.....:.....:.....:.....:
101 CGGTGCACAGCGCGGAACCTCGTGGCGCATCTGAATATCATCGGCGCAAT 52
499 eAsnValArgIlyProAlaSerIleArgProMetLysAsnIle 513
|||||.....:.....:.....:.....:.....:
51 AACGCTAAGAGTGCTCTGT.....:CCACATTTAATATC 18
```

seq_name: qb_est18:AA697364

seq_documentation_block:

LOCUS AA697364 786 bp mRNA EST 28-NOV-1998
DEFINITION HL02328 5prime HL Drosophila melanogaster head Bluescript
ACCESSION Drosophila melanogaster CDNA clone HL02328 5prime, mRNA sequence.
VERSION AA697364
KEYWORDS AA697364.1 GI:2700293
SOURCE EST.
ORGANISM fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscophora; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 786)

REFERENCE
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HMI Drosophila EST Project

JOURNAL Unpublished (1997)
COMMENT On Oct 30, 1997 this sequence version replaced gi:2160808.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
Plate: 23 row: C column: 4

High quality sequence stop: 687.
Location/Qualifiers

FEATURES
Source

1..786
/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="HL02328"

/clone_lib="HL Drosophila melanogaster head Bluescript"

/sex="male and female"

/dev_stage="adult"

/lab_host="SOLR"

/note="Organ: head-brain & sensory organ; Vector:
Bluescript SK; Site:1: EcoRI; Site:2: XhoI; Constructed
using Stragene ZAP-cDNA Synthesis kit. Oligo dT-primed
and directionally cloned at EcoRI and XhoI in Bluescript
SK(+/-)"

BASE COUNT 203 a 260 c 186 g 137 t
ORIGIN

alignment_scores:
Quality: 408.50 Length: 263
Ratio: 2.220 Gaps: 5
Percent Similarity: 69.962 Percent Identity: 37.643

alignment_block:

US-08-956-991-2 x AA697364 ..

Align seg 1/1 to: AA697364 from: 1 to: 786

```
1152 LeuAspGlyLeuGluLysTyrThrAsnTyrSerIleGlnValLeuAlaPh 1168
|||||.....:.....:.....:.....:.....:.....:
9 CTCATGCGCTTAAGACGATACCAACATCAATCAATCGACGCTCTGCC.. 56
1168 ethrArgAlaGlyAsp.GlyValArgSerGluGlnIlePheThrArgThr 1184
|||.....:.....:.....:.....:.....:.....:
57 .ACAACAGCGCGAAGAGGACGCTCCGACGCTCCATCCATCGCAACACC 105
1185 LysGluAspValProGlyProProAlaGlyValLysAlaAlaAlaLse 1201
:.....:.....:.....:.....:.....:.....:
106 GAACCCGATGTTCCGACACCACTGATGTAAGGCTCTGTATGAGG 155
1201 rAlaSerMetValPheValSerTrpLeuProProLeuLysLeuAsnGlyI 1218
:.....:.....:.....:.....:.....:.....:
156 TAAAGCGGCTATCTCTGATCTCTGCGCGCCGACACAGCAAGCAAGGCA 205
1218 lIleArgLysTyrThrValPheCysSerHisProTyr-ProThrValIle 1234
|||||.....:.....:.....:.....:.....:.....:
206 TTATCACCCAGTACACCGGTACTCCAAAGCGGAGCGGTGAGACTGAGAC 255
1235 SerGluPheGluAlaSerPro.AspSerPheSerTyrArgIleProAsnL 1251
:.....:.....:.....:.....:.....:.....:
256 AAGAGCCCAAGAGTTCCCACTACACATGAGTTTGAAGGCGACCGAGCA 305
1251 euserArgAsnArgGlnTyrSerValTrpValValAlaValThrSerAla 1267
|||||.....:.....:.....:.....:.....:.....:
306 TGGAGAGAGAACAGCCCTTGGAGTTCTGGGTACAGCTACGACCACTT 355
1268 GlyArgGlyAsnSerSerGluIleIleThrValGluProLeuAlaLysAl 1284
|||||.....:.....:.....:.....:.....:.....:
356 GGGAGACGACGACGACGATAGAGCATAGAGCCATGACGACGACGAGGTG 404
1284 apProAlaArgIleLeuThrPheSerGlyThrValThrThrProMetL 1301
|||||.....:.....:.....:.....:.....:.....:
405 .CCCGCC AGATCGCCTCTTGACGACACCTTCACTGACCACTTCAAGG 454
```

```

1301 ysaapilvalleuprocyslysalaValGlyaspProserProAlaVal 1317
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
455 AGACGCCAGATGCTGCTGCGCTGAGAGCCGCCAACCAGATC 504
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1318 LysTrpMetLysaspSeranGlyThrProSerLeuValThrIleAspGI 1334
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
505 ACATGG.....AAGATCAAGGGCGTGAA.....TTCAGTCCCAACA 542
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1334 YArgArgSerIlePheSeranGlySerPheIleIleArgThrValLysA 1351
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
543 TCGCATGCGGCTCTCCCGCAGCGATCGCTGATCAATCGGTAATC 592
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1351 IagLysaspSerGlyTYrTYrSerCysIleAlaAsnAsnAsnTrpGlySer 1367
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
593 GCCAGCATGCCGCGAGACTCTCGCCAGCCGCAAGCTGATTGCTAAG 642
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1368 AspGluIleIleLeuAsnLeuGlnValGlnValProProAspGlnProAr 1384
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
643 GACTTCATCAGCGCAACAGTgATTGTCTGGCACACACCAATCGCCCA 692
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1384 gLeuThrValSerIleThrThrSerSerIleThrLeuSerTrpLeup 1401
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
693 CGTCACCTCTCGGCACACCAATGTATGCCCTGACTGTAAAGTTGAAGC 742
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1401 rGIYAspAsnGlyGlySerSerIleArgGlyTYr 1412
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
743 CCCATGAGGAGACACTGTTCTCTGCATGATAC 777
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
seq_name: gb_gss3:B07180

seq_documentation_block:
LOCUS      B07180          465 bp          DNA          GSS          13-JUN-1996
DEFINITION CSRL72-h3-u CSRL flow sorted Chromosome 11 specific cosmid Homo
ACCESSION   B07180
VERSION     B07180.1  GI:1416458
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 465)
AUTHORS     Evans,G.A., Burbee,D., Davies,C., Hamner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S.,
            Harris,J., Deford,J., McFarland,J., Burzinski,K., Khan,M.,
            Kupfer,K. and Garner,H.R.
            Genomic Sequence Sampled Map of Chromosome 11
            Unpublished (1996)
            Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gervans@utsu.swned.edu, shane@mcdermott.swned.edu
PCR PRIMERs
FORWARD CTGTCAGATGAAGCACTC
BACKWARD TTCTTTAAATCCCGAC
Seq primer: 17
Class: cosmid ends
High quality sequence stop: 465.
Location/Qualifiers
      1..465
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="CSRL72-h3"
      /clone_lib="CSRL flow sorted Chromosome 11 specific"
      /sex="female"
      /cell-type="chimeric hamster somatic cell hybrid"
      /note="Vector: scos-1; Human Chromosome 11 specific cosmid
      library prepared from flow sorted human Chromosome 11

```

```

alignment_scores:
  Quality: 379.50      Length: 98
  Ratio: 4.170        Gaps: 2
  Percent Similarity: 92.857      Percent Identity: 78.571
alignment_block:
US-08-956-991-2 x B07180/rev
Align seg 1/1 to reverse of: B07180 from: 1 to: 465

1795 AsparGAlaArgSerSerMetValSerThrGluSerAlaSerThrTY 1811
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
429 GACAAAGGAGAGACAGCATGTGTCCACGTAGAGTCTTTCACCTA 380
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1811 rGluGluLeuAlaArgAlaTYrGluHisAlaLysMetGluGluGlnLeuA 1828
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
379 CGAGAGCTGCGCCCGCGCNTATGACATGCCAAGCTGAGAGAGAGCTGC 330
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1828 rGHisAlaLysPheThrIleThrGluCysPheIleSerAspThr.SerSe 1844
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
329 AGCAGCGCCAGATTGAGATCACCGAGTCTTCATCTCTGACAGTTTCNTC 280
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1844 rGluGlnLeuThrAlaGlyThrAsnGluTYrThrAspSerLeuThrSer. 1860
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
279 TGACCGAG.TGACCAAGGACACACAGAGAACGCCGACAGCATGACATCA 230
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1861 ..SerThrProSerGluSerGlyIleCysArgPheThrAlaSerProPro 1876
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
229 TGAGCAGACACNTCAGAGCNTGCATCTGCCGNTTTACCGCTCACCAACC 180
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1877 LysProGlnAspGlyGlyArgValMetAsnMetAlaValPro 1890
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
179 AAGCCCGAGATGCGGACCGGCCAAACGCTGCTGCCC 138
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
seq_name: gb_est27:AI404957

seq_documentation_block:
LOCUS      AI404957          569 bp          mRNA          EST          08-FEB-1999
DEFINITION GH24836.5prime GH Drosophila melanogaster head P012 Drosophila
ACCESSION   AI404957
VERSION     AI404957.1  GI:4248044
KEYWORDS    EST.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 569)
AUTHORS     Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Broksstein,P., Lewis,S. and Rubin,G.M.
            BDGP/HMI Drosophila EST Project
            Unpublished (1997)
            On Apr 21, 1998 this sequence version replaced gi:3073067.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of vector sequence at both ends, this
            sequence has been determined to be the complete cDNA insert.
            Plate: 248 row: C column: 12
            High quality sequence stop: 560.
            Location/Qualifiers
      1..569
      /organism="Drosophila melanogaster"

```

```

/db_xref="taxon:7227"
/clone="GH24836"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: pot2; Site:1: EcoRI; Site:2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
BASE COUNT      145 a      150 c      154 g      120 t
ORIGIN

```

```

alignment_scores:
  Quality: 355.50      Length: 185
  Ratio: 2.799      Gaps: 4
  Percent Similarity: 68.649      Percent Identity: 40.541

```

```

alignment_block:
  US-08-956-991-2 x A1404957

```

```

Align seg 1/1 to: A1404957 from: 1 to: 569

```

```

1145 ThThrcInProSerLeuGluLeuAspGlyLeuGlyLysTyrThrAsnTy 1161
      ||||| ..... ||| |||||:|||||
      7 ACACACACACTCACCATGCTGTCACAGACTTCGCAATATACCAACTA 56
1161 rSerIleGlnValLeuAlaPheThrArgAlaGlyAspGlyValArgSerg 1178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      57 CAGATTGAGGTTTGGCCACACAGCGCTATGGCGAGAGGTGTATCGA 106
1178 LucInIlePheThrArgThrLysGluAspValProGlyProAlaGly 1194
      :: ||||| ::||| ::||| ::||| ::||| ::||| ::|||
      107 AGCCATGTTTTCACACAGGAGAGATGTGCTGAGCGCTCGGCAGAT 156
1195 ValLysAlaAlaAlaSerAlaSerAlaSerMetValPheValSerPheLeuP 1211
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
      157 ATCAAGTTGTGTCACACTATCTCAATCACTATATCTTGTGTTGCC 206
1211 OPLeuLysLeuAsnGlyIleIleArgLysTyrThrValPheCysSerH 1228
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
      207 ACCCAATGACCGACGCTGTGATCACAAGTACAGTCTT..... 246
1228 IsProTyrProThrValIle.....SergLeuPhe 1237
      ||| ||||| ::||| ::|||
      247 .....TACACTCGCTAGTGAACGTCGCGAGAACTGAACAATGAGAA 291
1238 GluAlaSerProAspSerPheSer...TyrArgIleProAsnLeuSerAr 1253
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
      292 CGTAGTCTGCATCACAGCAGCGGTATTACGAAGCCAAAGAGTGCATCC 341
1253 GAsnATGgIntTyrSerValTTrpValAlaValAlaThrSerAlaGlyArgG 1270
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
      342 GCATATGAGGATATCACTGTGCTGACGGCCAGCACTGATGGGGGAGG 391
1270 LysAsnSerSerGluIleIleThrValGluProLeuAlaLysAlaProAla 1286
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
      392 GCAGAGTTCCCGTGTTCCTTCACAAATACCAAGATTCGATGGGG 441
1287 ArgIleLeuThrPheSerGlyThrValThrThrProTyrPheLysAsp11 1303
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
      442 AGGATCATATCTTGGTGGTCCAGTGTGACACGCTTGGCATCCACCGT 491
1303 eValLeuProCysLysAlaValAlaGlyAspProSerProAlaValLysTTPM 1320
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
      492 TACTTACCTGACGGCGGTGGCAAAACCAAGCC.....GAGTGT 535
1320 eLys 1321
      ::|||
      536 TCAA 540

```

```

seq_name: gb_gss1:FR0021997

```

```

seq documentation block:
LOCUS      FR0021997      610 bp      DNA
DEFINITION F. rubripes GSS sequence, clone 070J16AB10, genomic survey sequence.
ACCESSION AL014868
VERSION   AL014868.1 GI:2681236
KEYWORDS  GSS: genome survey sequence.
SOURCE    Fugu rubripes.
ORGANISM  Fugu rubripes.

```

```

REFERENCE  Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
AUTHORS    Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.
            1 (bases 1 to 610)
            Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umraniya, Y.,
            Williams, G. and Brenner, S.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk
COMMENT    Vector: pBluescript II KS
            V.type: phagemid
            PRIMER: KS

```

```

DESCR      One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.

```

```

FEATURES   location/Qualifiers
            source

```

```

            1..610
               /organism="Fugu rubripes"
               /db_xref="taxon:31033"
               /clone_lib="cosmid 070J16"
               /clone="070J16AB10"
BASE COUNT  158 a      133 c      150 g      147 t      22 others
ORIGIN

```

```

alignment_scores:
  Quality: 346.50      Length: 187
  Ratio: 3.013      Gaps: 7
  Percent Similarity: 61.497      Percent Identity: 43.850

```

```

alignment_block:
  US-08-956-991-2 x FR0021997/rev

```

```

Align seg 1/1 to reverse of: FR0021997 from: 1 to: 610

```

```

586 SerGlnSerValHisValThrValLysValProPheIleGlnProPh 602
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
      555 TCTTCTTCTTCCACTTA.....CTGCCACCTTCATCCACACNTT 515
602 eGluPheProArgPheSerIleGlyGlnArgValPheIleProCysValV 619
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
      514 NGATATCCNTTNTACNTCAGTGGGAGAGATCATTTACATCGCTGG 465
619 aValSerGlyAspLeuProIleThrIleThrTrpGlnLysAspGlyAr 635
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
      464 TGTATNTGGGACATGCCATAGCATCACTGCGCAACAAATGCGCA 415
635 gProIle...ProGlySer...Leu.GlyValThrIleAspAsnIleAsp 649
      ||||| ||| ||| ::||| ::||| ::||| ::|||
      414 GCCCATTTGCCAGNCNTCCGNTTCTGGCTGGCAATAGAGACCAAGAG 365
649 PheThrSerSerLeuArgIleSerAsnLeuSerLeuMetHisAsnGlyAs 666
      ||| ||||| ::||| ::||| ::||| ::||| ::|||
      364 TTCATGAGTCTTNTGCAAAATCACCAGGTGACGTTAAAGCAACGAA 315
666 nTyrThrCysIleAlaIleArgAsnGlnAlaIleAlaValGlnHisGlnSerG 683
      ||||| ||||| ||| ||||| ::||| ::||| ::|||
      314 CTACAGTCATGCCGCCAGCAACAGCAGCAGCTGTACGTTGGAGCGAG 265
683 InLeuIleVal..Arg.....
      ||||| ||||| |||
      264 AGCGATCGTTACAGGTGAGGAGGATGATTACACCAAGAGTGGACAGAA 215
687 .....

```

214 TAGTCGGCTCAAAATTTGGTATCTACCCAGAGAGACCCNATNNATAAG 165
688 ValProPolysph 692
164 TATTTCCTAGCTTATACATCTGCTGTTTTCAGTCCGCGGCTT 115
692 eValValGlnProArGAspGlnAspGly1LeTyrGlyLysAlaVal1LeI 709
114 TGTACTGAGCCCAACATCAAGACTGATTATGSCAAAGCTGGAGTGTG 65
709 euAsnCysSerAlaGly1TyrProValProThr1LeVal1TrpLysPhe 725
64 TCAACTGTTNCTGGAGGGTACCTCCCCCAAGATCATGTGGAACAC 15
726 SerLys 727
14 GCCAA 9
seq_name: gb_est128:A1542894
seq_documentation_block: 696 bp mRNA EST 22-MAR-1999
LOCUS A1542894
DEFINITION SD09407:5prime SD Drosophila melanogaster Schneider L2 cell culture
pot12 Drosophila melanogaster cDNA clone SD09407 5prime, mRNA
sequence.
ACCESSION A1542894.1 GI:4460267
VERSION A1542894
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscophora; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 696)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Broksstein, P., Lewis, S. and Rubin, G.M.
BDGP/HMMI Drosophila EST project
Unpublished (1997)
On Mar 16, 1998 this sequence version replaced gi:2961848.
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 94 row: A column: 7
High quality sequence stop: 558.
Location/Qualifiers
FEATURES
Source
1. 696
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD09407"
/culture_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot12"
/lab_host="DH5-alpha"
/note="Vector: pot12; Site_1: EcoRI; Site_2: XhoI. Sized
fractionated cDNAs were directly ligated into pot12.
Plasmid cDNA library."
BASE COUNT 183 a 197 c 183 g 133 t
ORIGIN
alignment_scores: Length: 230
Quality: 345.50 Gaps: 3
Ratio: 2.273
Percent Similarity: 66.087 Percent Identity: 31.304
1. block:
56-991-2 x A1542894 ..
1/1 to: A1542894 from: 1 to: 696
rgAspGlnAspGly1LeTyrGlyLysAlaVal1LeI euAsnCys5c 712

111
1 CCCACCGATAGGACCTTCGCCAGGATCCGATGCCAAGGTGATGCA 50
712 rAlaGly1TyrProValProThr1LeVal1TrpLysPheSerLysGlyA 729
51 GGCTGATGGCTTCCCAACCCCAAGTTACATGGAAGAACGATGGCG 100
729 laGlyValProGlnPheGlnPro1LeAlaLeuAsnG1Arg1LeGlnVal 745
101 ACACCCCGAGAGTACAGATCTAAGACAGACGACAAACATCCGCGT 150
746 LeuSerAsnGlySerLeuLeuLeuLysHisValValGluGluAspSerG 762
151 ... GAAGAGGGGACCTCGCATGTCGACAAACATCCAAAGACCAAGAGG 197
762 yTyr1TyrLeuGlyLysValSerAsnAspValGlyAlaAspValSerLys 779
198 CTACTATCTGTGGAGGCTATCATGATAGCTCCGCGCTCTCGCTG 247
779 ermety*LeuThrValLys1LeProAlaMet1LeThrSerTyrProAsn 795
248 TGATTATGATCAAGCTTCAGGCCCTCCAGAGTTCAAGAGAACTGCGC 297
796 ThrThrLeuAlaThrGlnGlyGlnLysLysGluMetSerCysThrAla1 812
298 AACCAAGACCGCCGACGAGAGAACCCGCGTACTCCAGTGGAGCGCA 347
812 sGlyGlyLysPro1Le1LeValArg1TrpGlyLysGluAspArg1Le1Lea 829
348 GGGCGGAAGCCCATTTGGCATCTTGTGGAAACATGAACAACATGCGAC 397
829 snProLysMet...AlaArg1TyrLeuValSerThrLysGluValGlyL 844
398 ACCCAAGAACGACATCGGTACCATTTCTGAGAGAACTCTTCCACC 447
845 GluVal1LeSerThrLeuG 1LeuProThrValArgGluAs SerG1 861
448 GCAGTATATGTCAGTGTGATGCAAGCCGACGAGAGATCGCATCCGC 497
861 yPhePheSerCysHisAla1LeAsnSerTyrGlyLysAspArg1Le1 878
498 CCAATTCACCTGTGTGCGACCAATATGCTTGATGCGATCGACGACCA 547
876 leGlnLeuThrValGlnGluProProAspProPro...Glu1LeGlu1Le 893
548 TAAATATGATTGTCACGAACTTCGCGAAGCATATGCTTTGAGGTA 597
894 LysAspValLysAlaArgThr1LeThrLeuArgTyrThrMetGlyPheAs 910
598 CTCGACAAATCCGAGAGCTTCCTGCAGCTGAGCTGGGGCAACCTTACGA 647
910 pGlyAsnSerPro1LeThrGly1TyrAsp1LeGluCysLys 923
648 TGGCAACTCCCTCTGGACAGGTACATCATGTAGTTAAG 687
seq_name: gb_est137:AW014717
seq_documentation_block: 409 bp mRNA EST 10-SEP-1999
LOCUS AW014717
DEFINITION UT-H-B10-aae-a-07-0-UT-s1 NCI-CGAP-SubI Homo sapiens cDNA clone
IMAGE:2709012 3', mRNA sequence.
ACCESSION AW014717
VERSION AW014717.1 GI:5863474
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT

On May 18, 1998 this sequence version replaced gi:1138023.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/ILM at:
www.bio.11n1.gov/db/tp/image/image.html The following repetitive
elements were found in this cDNA sequence: 1-21,
>AT richlow complexity
Seq primer: M13 Forward
POLY-A-yes.

FEATURES

Source
Location/Qualifiers
1..409
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2709012"
/clone_1b="NCI CGAP Sub1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Ph rmacia) with a modified
polylinker. Site 1: Not I; St e_2: Eco RI; The
NCI CGAP Sub1 library is a subtracted library derived from
BI. BI constitutes a mixture of 21 normalized or
subtracted NCI CGAP libraries: NCI CGAP Co4,
NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,
NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
NCI CGAP Br23, NCI CGAP Co8, NCI CGAP CUL1, NCI CGAP Lel2,
NCI CGAP Br23, NCI CGAP Lu5, NCI CGAP Lu24,
NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6, NCI CGAP Brn25.
These 21 libraries were pooled and a single-stranded DNA
preparation of the resulting mixture was used as a tracer
in a subtractive hybridization with a driver whose
composition is detailed below: NCI CGAP Kid3 pool 1 LLM
3334-3337, 3682-3683, 3798-3803 (IMAGE Cloneds
132376-132391, 1456008-1456775, 1500552-1502855)
NCI CGAP Kid5 pool 1 LLM 3338-3342, 3722-3725, 3776-3778
(IMAGE Cloneds 1323912-1325831, 1471368-1472903,
1492104-1493253) NCI CGAP Lu5 pool 1 LLM 3575-3582,
3851-3854 (IMAGE Cloneds 1414920-1417991,
1520904-1522439) NCI CGAP GC4 pool 1 LLM 3164-3167,
3716-3720, 3733-3735 (IMAGE Cloneds 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI CGAP Pr22 pool 1
LLM 2457-2459, 2758-2759, 3062, 368 (IMAGE Cloneds
985608-986739, 1101192-1101939, 1217928-1220615)
NCI CGAP Co10 pool 1 LLM 2644-2653, 2871-2872 (IMAGE
Cloneds 1057416-1061255, 1144584-1145351) The resulting
subtracted library contained 530,000 recombinants.
Subtraction was performed as previously described
[Ponaldi, Lennon & Soares (1996): Normalization and
Subtraction: Two Approaches To Facilitate Gene Discovery.
Genome Research 6, 791-806.
TAG LIB-NCI CGAP Kid3
TAG TISSUE-Kidney
TAG_SFO-AATGC"

BASE COUNT 72 a 110 c 121 g 106 t
ORIGIN

alignment_scores:

Quality: 344.00 Length: 100
Ratio: 3.954 Gaps: 0
Percent Similarity: 87.000 Percent Identity: 60.000

alignment_block:

US-08-956-991-2 x AW014717/rev

Align seg 1/1 to reverse of: AW014717 from: 1 to: 409

404 AspGlyThrProLysIleIleSerAlaPheSerGluLysValValSerPr 420

|||||
331 GATGGCAGCCCGCATCTCTGCTCTCAACGAGAGTGTCAACC 282
420 OAlaGluProValSerLeuMetCysAsnValLysGlyThrProLeuProT 437
281 CGGGAGCAGATTCTCTCATGTGTGCGCCCAAGGCGCCCGCCCA 232
437 hTrleThrPThrIleAspAspProIleLeuLysGlySerHis 453
231 CGGTCTCTGGGCGCTCGACGATGACCCATGTGCGGATGCGACCC 182
454 ArgIleSerGlnMetIleThrSerGluLysAsnValValSerTyrLeuAs 470
181 CGCACCACAGTACACAT""CGACGGCACCACCATCAGCCACATGMA 132
470 nIleSerSerSerGlnValArgAspGlyValTyrArgCysThrAla 487
131 CGTCACAGCCCGCCAGATCCGCGCGGGCGGTACCGGTCCACAGCC 82
487 snAsnSerAlaGlyValValLeuTyrGlnAlaArgIleAsnValArgGly 503
81 GGAAGTGTGTGGGAGTCTGATATACAGCCGAGATATAAGTAGAGT 32

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:59:04 ; Search time 4987.54 seconds
(without alignments)
4854.765 Million cell updates/sec

Title: US-08-956-991-10
Perfect score: 6413
Sequence: 1 tgaactgagcgagcgacg9.....gaaattgccaaaataatt 6413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	qb_est1:*
21:	qb_est2:*
22:	qb_est3:*
23:	qb_est4:*
24:	qb_est5:*
25:	qb_est6:*
26:	qb_est7:*
27:	qb_est8:*
28:	qb_est9:*
29:	qb_est10:*
30:	qb_est11:*
31:	qb_est12:*
32:	qb_est13:*
33:	qb_est14:*
34:	qb_est15:*
35:	qb_est16:*
36:	qb_est17:*
37:	qb_est18:*
38:	qb_est19:*
39:	qb_est20:*
40:	qb_est21:*
41:	qb_est22:*
42:	qb_est23:*
43:	qb_est24:*
44:	qb_est25:*

45:	qb_est26:*
46:	qb_est27:*
47:	qb_est28:*
48:	qb_est29:*
49:	qb_est30:*
50:	qb_est31:*
51:	qb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	qb_est33:*
60:	qb_est34:*
61:	qb_est35:*
62:	qb_est36:*
63:	qb_est37:*
64:	qb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	qb_est39:*
70:	qb_est40:*
71:	qb_est41:*
72:	qb_est42:*
73:	qb_est43:*
74:	qb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	qb_gss1:*
80:	qb_gss2:*
81:	qb_gss3:*
82:	qb_gss4:*
83:	em_gss1:*
84:	em_gss2:*
85:	em_gss3:*
86:	em_gss4:*
87:	qb_gss5:*
88:	qb_gss6:*
89:	qb_gss7:*
90:	qb_gss8:*
91:	qb_gss9:*
92:	em_gss5:*
93:	em_gss6:*
94:	em_gss7:*
95:	em_gss8:*
96:	em_gss9:*
97:	em_gss10:*
98:	em_gss11:*
99:	qb_gss10:*
100:	qb_gss11:*
101:	em_gss12:*
102:	qb_gss12:*
103:	qb_gss13:*
104:	qb_gss14:*
105:	qb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	415.8	6.5	439	43	AT243628 qb88a03.x
C 2	399.2	6.2	541	64	AW046213 UT-M-BH1-

C	3	397.4	5.2	419	42	A1094516	A1094516	0Y63f311.S
C	4	374.2	5.8	510	25	N80326	N80326	Y291f03.r1
C	5	358.8	5.6	430	25	M64532	M64532	Y291f03.s1
C	6	343.8	5.4	355	37	AJ003472	AJ003472	
C	7	341.2	5.3	458	46	A14474704	A14474704	UT-R-BT0-
C	8	318	4.7	3021	20	Z41519	Z41519	HSCXW042.N
C	9	303.4	5.0	321	41	A10511251	A10511251	OY49e604.x
C	10	271.2	4.2	427	21	AM159035	AM159035	Z450e05.x
C	11	265	4.1	310	21	F13426	F13426	HSCXAO21.N
C	12	262.4	4.1	365	64	AM048531	AM048531	UT-M-BH1-
C	13	260.4	4.1	294	20	Z45894	Z45894	HSCXW041.N
C	14	254.8	4.0	337	29	AA170935	AA170935	MS45M05.r
C	15	242.4	3.8	252	21	F09100	F09100	HSCXAO22.N
C	16	213.4	3.3	264	46	A1448860	A1448860	MS45M05.x
C	17	187	2.9	187	39	AA878733	AA878733	Oe0d001.s
C	18	175.6	2.7	619	79	FR0002941	FR0002941	F.rubripes
C	19	163.2	2.5	402	62	AM048129	AM048129	UT-M-BH1-
C	20	156	2.4	156	64	A1880033	A1880033	ap24b07.x
C	21	147.4	2.3	409	63	AM014717	AM014717	UT-H-BT0-
C	22	143	2.2	147	79	HSMC18B12	X88325	H.sapiens.D
C	23	142.4	2.2	465	81	B07180	CSRU72-h3-u	
C	24	138.4	2.2	450	47	A1498283	A1498283	tmLc111.x
C	25	123	1.9	646	42	A1106884	A1106884	GMO6134.S
C	26	122.8	1.9	602	79	FR0022038	AL014909	F.rubripes
C	27	120.6	1.9	422	79	FR0030733	AL027102	Fugu rubr
C	28	118.4	1.8	437	37	FR0002934	Z86717	F.rubripes
C	29	113.4	1.8	298	31	AA335603	AA335603	EST28915
C	30	111.4	1.7	315	104	AO522303	AO522303	RPCT-11-
C	31	105.4	1.6	935	79	CNS006KX	AL066051	Drosophil
C	32	104.6	1.6	282	71	AV264554	AV264554	AV264554
C	33	104	1.6	932	79	CNS007QD	AL066742	Drosophil
C	34	98.6	1.5	776	79	CNS010KX	AL099332	Drosophil
C	35	97.4	1.5	935	79	CNS006KX	AL069051	Drosophil
C	36	95	1.5	1101	80	CNS01523	AL104901	Drosophil
C	37	94.8	1.5	1080	88	AO8933056	AO8933056	H5_4832_A
C	38	94	1.5	1101	80	CNS0153F	AL104949	Drosophil
C	39	93.6	1.5	925	79	CNS0091P	AL050312	Drosophil
C	40	93.6	1.5	1035	80	CNS01760	AL08162	Drosophil
C	41	92.8	1.4	99	79	HSMC34009	X88135	H.sapiens.D
C	42	92.6	1.4	925	79	CNS0091P	AL053013	Drosophil
C	43	92.2	1.4	544	79	FR0022080	AL014951	F.rubripes
C	44	91.8	1.4	932	79	CNS007QD	AL066742	Drosophil
C	45	91.4	1.4	894	88	AO893147	AO893147	HS-3131.B

ALIGNMENTS

```

RESULT      1
LOCUS       AI243628/c
DEFINITION  AI243628      439 bp      mRNA      EST      01-DEC-1998
            gi888403.x1 Soares_NF_L_T_GBC_S1 Homo sapiens cDNA clone
IMAGE       1854028 3', mRNA sequence.
AI243628
ACCESSION   AI243628
VERSION
KEYWORDS
SOURCE
ORGANISM    human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 439)
            NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
            Tumor Gene Index
            Unpublished (1997)
JOURNAL
COMMENT      On Jan 17, 1998 this sequence version replaced gi:2043430.
            Contact: Robert Strausberg, Ph.D.
            Tel.: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.lnl.gov) for further information.
            Insert Length: 476 Std Error: 0.00
            Seq primer: -400p from Gibco.

```

FEATURES		Location/Qualifiers
source		1. 439 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1854028" /clone.lib="Soares_NFL_T_GBC_S1" /lab_host="DH10B" /note="Organ: pooled. Vector: pRTD-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung Nhd19w, testis NHT, and B-cell NCI-G46-GB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT		86 a 104 c 115 g 133 t 1 others
ORIGIN		
Query Match	6.5%; Score 415.8; DB 43; Length 439;	
Best Local Similarity	98.1%; Pred. No. 1.2e-08;	
Matches 420;	Conservative 0; Mismatches 8; Indels 0; Gaps 0;	
QY 5986	tttctgtaaccgagatgtgtcagacaccacagcagagctgaagccttagacaagctgct	6045
Db 432	TTTTTTGTTAAACCGAGGTGCTCCAGGCACCCACANGGACCTGACCTTAGACAAAGATGCT	373
QY 6046	tggagacc cagaaaagccggaacctgaaagcccccacggtctcgtgagccatcccgatgg	6105
Db 372	TGGAAACC.AAGAAAGCCGGAGACCTGAAAGCCGCCACGAGCTCTGGAGCCATCCGATGG	313
QY 6106	aagccgccttcctccctccctcccaagagagaagaaagcagctgtgtgacagccggggccgtgg	6165
Db 312	AAGCCGCTCTCTCTCCCTCTCTCCACGAGAGAAAGACAGTGTGTGGACCGGGGCCGTGG	253
QY 6166	ccccattaccctcagcggagagagaagagcgtcgagacagcagcttaaatgaacagctccc	6225
Db 252	CCACATTACTCAGCGGGAGGGAGACAGAGCTGGGACAGCACTTAATAATGACACACTCCC	193
QY 6226	aagaatcactgctcgaactccgqgqccalttgaagaagaaacaatccttaagcaaaatctt	6285
Db 192	AAGAAATCACTGCTCCACACTCCGGCGGCCATTGGAAGGAACAATCCTTAGCAAAATCTT	133
QY 6286	acaccctgtgtataacagacagcaatgactcgaagacagcgtgttaataacaatttaacaact	6345
Db 132	ACACCTGTGATTAACAGACAGCATGACTGCAACGCGGTTTAAATTCATTCACAACTT	73
QY 6346	caatcaaaagctaccccttttttlaagyaattccaattataatlaataaagaaattggcaa	6405
Db 72	CAATCAAAAGCTACCTTTTTTTACGGAATTCGAATTTATTTAATTAAGAAAAATTGCCAA	13
QY 6406	aataatatt 6413	
Db 12	AATTAATT 5	
RESULT 2		
LOCUS	AM046213	
DEFINITION	UT-M-BH1-akw-g-12-0-UT.s1 NIH-BMAP_M.S2 Mus musculus cDNA clone	
ACCESSION	AM046213	
VERSION	AM046213.1	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus.	
REFERENCE	1 (bases 1 to 541)	

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jun 5, 1998 this sequence version replaced g1:3188463.
 Contact: Chih, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestrail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized hypothalamus library cDNA library preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence:
 1-52, >At_1rch#Low_complexity
 Seq primer: M13 Forward
 POLY-A-yes.

FEATURES Location/Qualifiers
SOURCE 1. 541
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U1-M-BH1-akw-g-12-0-U1"
 /clone_1lb="NIH_BMAP_M_S2"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_M_S2 library is a subtracted library derived from NIH_BMAP_M_S1, which in turn is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH_BMAP_M_S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries.
 TAG_L1b=NIH_BMAP_M_S2
 TAG_TISSUE=hypothalamus
 TAG_SEQ=CAGTA"

BASE COUNT 105 a 127 c 144 g 165 t
ORIGIN

Query Match 6.2% Score 399.2; DB 64; Length 541;
 Best Local Similarity 88.9% Pred. No. 1.7e-65;
 Matches 466; Conservative 0; Mismatches 53; Indels 5; Gaps 3;

OY 5895 tcagatggaagaagatgaatgcaatgcaaaagga-atcgagccaggtgacct 5953
 ||||||| ||||| ||||| ||||||| ||||||| ||||||| |||||||
 Db 541 TCAGGATGGAGGAGGTGTGAACATGGGGTTCCAAAGGCCCATCGCGCAGCT 482
 OY 5954 catatctggcccatctacatagatgagctttgttaaacgaggtggtccagcac 6013
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 481 CATACACTGCTCCCTACTACTAGCAATGAGCTCTTGTTAAACGGCGGCACACGAC 422
 OY 6014 cagcagagacctgagcttaagacaacatgcttggaaacctcaaaagccgagacctgaa 6073
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 421 CAGAGAGGACCTGAGTTTAGACAAGCGTCTTGGAACCCAGAAAAAGTGGACCTTGAA 362
 OY 6074 ggcgcccaagctctgagaccctcccgatggaagcgcgctcctccgctccacgag 6133
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 361 AGGCCACAGGTCCTGTAGACCCACCTATGAGAGGCTCTCTTCACCTTTCCACGC 302
 OY 6134 agaaga---cagtcgagcgccgagggccgtggccacatctaccagcgaggaagc 6190
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 301 AGAAGACAGCAGTCGTGGCAACAGGGGCTGTGGCCACTTACTACGACAGAGGCTGC 242
 OY 6191 agagctggagcagcagcttaaatgagcagctcccaagaatctgtcgtacctccgag 6250
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 241 AGAGCTGGACAGCGACGCTTAATATGACAGCTCCCAAGAGTACTGCTGCCTCCGGG 182
 OY 6251 ccattgaaaggaacaatcctctacgcaaatcttaccacctgtgtataacagaaagca-t 6309
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 181 CCATTGGAAGGAACAATCCCTCCGCAAAATCTTACCTGTGATATACATAGCACC 122
 OY 6310 gactgagcagcggtgtgtaataatcaatcaatcaatcaatcaatcaatcaatcaatcaat 6369
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 121 TGATGACAGCGGTGTGTAATATCAATTTAAACGAGCCATCAAGCTACCTTTTATAC 62
 OY 6370 ggaattccaatatataatataaagaagaattgccaaatatatt 6413
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 61 TGAATCCGATATTATATTAAAGAAATGGCCAAATATATT 18

RESULT 3
LOCUS AI094516/c 419 bp mRNA EST 24-SEP-1998
DEFINITION oy63f11.s1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1670541 3', mRNA sequence.
ACCESSION AI094516
VERSION AI094516.1 GI:3433492
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 419)
AUTHORS NCI/NIHNS-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index
COMMENT Unpublished (1998)
 On Jan 19, 1998 this sequence version replaced g1:2284845.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.L.Strausberg@nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html

Insert Length: 659 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amersham.

FEATURES Location/Qualifiers
SOURCE 1. 419

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1670541"
 /clone_1lb="NCI_CGAP_Brn23"
 /rissue_type="glioblastoma (pooled)"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACCAATCTGAGAGGAGCGCGCATATCTTTTGTTTTGTTTT T 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library is normalized, and was constructed by Bento

|||||
RESULT 7
LOCUS A1454704 458 bp mRNA EST 05-JUL-1999
A1454704/c
DEFINITION UI-R-BT0-qk-h-03-0-UI-s1 UI-R-BT0 Rattus norvegicus cDNA clone
A1454704
UI-R-BT0-qk-h-03-0-UI 3', mRNA sequence.
A1454704
VERSION A1454704.1 GI:4295587
KEYWORDS EST.
SOURCE Norway rat.
Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 458)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
REFERENCE Normalization and subtraction: two approaches to facilitate gene
AUTHORS discovery
TITLE Genome Res. 6 (9), 791-806 (1996)
JOURNAL 97044477
MEDLINE
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980496.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics (www.resgen.com) This clone is also
available through the I.M.A.G.E. Consortium at LNL
(info@image.llnl.gov). IMAGE ID=1788947
Seq primer: M13 Forward
POLY-A-No.

FEATURES
source
1..458
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT0-qk-h-03-0-UI"
/clone_lib="UI-R-BT0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73b-pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; This library
(UI-R-BT0) consists of a mixture of individually tagged
normalized libraries constructed from rat hippocampus,
thalamus, mid-brain, medulla, corpus striatum, cerebral
cortex and testis. The tag used to identify the source
tissue is a string of 3-6 nucleotides present between the
Not I site and the oligo-dt track which allows
identification of the library of origin of a clone within
the mixture. This library was then subcloned using a
driver consisting of a mixture of all clones from UI-R-A0,
UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and
UI-R-C2p."

BASE COUNT 107 a 112 c 123 g 116 t
ORIGIN

Query Match 5.3% Score 341.2; DB 46; Length 458;
Best Local Similarity 84.1%; Pred. No. 14e-54;
Matches 385; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 2024 ctctcgtgattgactcgtattactcattacgtatgaagaactcaactgct 2083
|||||
DB 458 ctctcgtgattgactcgtattactcattacgtatgaagaactcaactgct 399
QY 2084 tctcttaaccaccgccaagctgcatcttgagaacatgga tctcttaacttcaagatgt 2143

|||||
Db 398 TCTTTTCAACCCAGCAGAGTGGCGTTTGAGAACATGTGACTTGAACTCTCAGATGT 339
QY 2144 gcaaaaggaagtggagagggaagtaacacgtcgaacgtgtgtgtcaaccacactctc 2203
|||||
Db 338 GCAGAAAGAGTTGACGAGGAGAGTACATGTAGTGGTGGTGCACCCAGCTCTC 279
QY 2204 caccagccagagcgtcgaacgtgaacgtgaagttccgccttcatacaacccttgaatt 2263
|||||
Db 278 CACGAGCAGAGTGTCCAGCTGACGTGACATCAAGTGTGCTTTTATTCACACCTTTGAGTT 219
QY 2264 tccaagattccattggcgaagcgggtcttcacccctgtgtgtgtcgaaggaactt 2323
|||||
Db 218 CCGAAGATTATATGCGTACAGCGGTATTCATCCCATGTGTGTGCTCGGGGACTT 159
QY 2324 acccatcagatcactcgtgcagaagaatgacgcgcacatccctcgtggagccttgggtgac 2383
|||||
Db 158 ACCCATCACCATACGTGGCAGAGAGATCCCTCCGATCCACAGAGCTCGGTGTAC 99
QY 2384 catgacaatatgactcaacgagctcctcttgaggattccaatctctcgtcatgacaa 2443
|||||
Db 98 CATCGACACATCGACTTCACAGCTCTTGAGAGATCTCCACCTCTCAGTATGCACA 39
QY 2444 tgggaattacactgcatagcccggaatgaagccgcg 2481
|||||
Db 38 CGGAATTAACATGATCGAGCGAAGAGCGGCGG 1

RESULT 8
241519/c
LOCUS 241519 321 bp mRNA EST 10-NOV-1994
DEFINITION HSC2WH042 normalized infant brain cDNA Homo sapiens cDNA clone
c-zwh04 3', mRNA sequence.
241519
ACCESSION 241519
VERSION 241519.1 GI:567667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 321)
REFERENCE Aufray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pletu,G., Pouliot,Y.,
Sebastiani-Kabakchis,C. and Tessier,A.
IMAG: molecular integration of the analysis of the human genome
and its expression
C.R.Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
JOURNAL Contact: Genethon
MEDLINE Genexpress-Genethon
COMMENT Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read. 21 r removed at sequence 5' end
Genexpress library_idt: C; Genexpress_sequence_idt: alc-zwh04
Seq primer: (-21)M13 universal.

FEATURES
source
1..321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-zwh04"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: Jafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate=muscular atrophy patient; tissue=total
brain; total mRNA was oligo-(dt) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the

Fax: 516 367 8874
Email: mcomble@cshl.org
Plate: za50 row: e column: 05
Seq primer: M13 universal forward primer
High quality sequence stop: 427.
Location/Qualifiers

FEATURES

source
1. 427
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="za50e05"
/clone_lib="Xenopus EST library"
/tissue_type="total brain tissue"
/cell_line="W2-TGA"
/dev_stage="laddpole"
/note="Vector: Lambda zap I; Site-1: XbaI; This library was supplied by Holly Cline (Cold Spring Harbor Labs).
cDNA synthesis with oligo df Xba I (Xba I cloning site).
RNA: stage 50-56 tadpoles, total brain tissue, GTC extraction method."
BASE COUNT 114 a 122 c 91 g 100 t
ORIGIN

Query Match 4.2% Score 271.2; DB 71; Length 427;
Best Local Similarity 79.8% Pred. No. 2e-41; Mismatches 78; Indels 7; Gaps 1;
Matches 335; Conservative 0;

Oy 1678 ttaattctacacctgaagaaagtgtagtcacagacagacggttcctt..vgtgca 1737
15 TCATCTCTGCATTAGCGAAGAGTGTGACGCCGAGAACCTGTCTCTATGTGCA 74
Oy 1738 acgtaagaaagacaccttgcccacgacgtgagacctggagatgaccgatttcca 1797
75 ACGTAAAGGAAGCGCTCTTCGACGATCCTTGAGCGCTGGACGACGCCCATATCA 134
Oy 1798 aggggtgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1857
135 AGACGGCCACACAGATCAATAGTCACTGAGGCGACGCTGTGACGTTAC 194
Oy 1858 tgaacatctcagctccagctccagctccagctccagctccagctccagctccagctcc 1917
195 TGAACATCACCAACTCAAGTCCGATGCGGGGTGTACCGATGACGCGTAAACACT 254
Oy 1918 cggcgagagctctcgtcgtcagctcgaataacgtaagagagcctcaagcattcgac 1977
255 CTGGGGCGCTGCTTGTTCACAGGCTCGAATAACCT-----CCTGGAATATCCGTC 307
Oy 1978 caatgaaacacacacagacagacagacagacacacacacacacacacacacacacac 2037
308 CAAATGAAACATCAGCGCATAGTGGCGGACACATACATCCATCCCGCTCATTTG 367
Oy 2038 gctatcgtatttaccatataatgtagtaagaacttaacctgcttcttcaaccacc 2097
368 GATATCCCTACTACTCATCAAGTGTACAAACACTTAACCTCCTCCATTTAATCACC 427
Db

RESULT 11
F13426 310 bp mRNA EST 15-MAR-1995
LOCUS HSC2A021 normalized infant brain cDNA Homo sapiens cDNA clone
DEFINITION C-2x802, mRNA sequence.
ACCESSION F13426
VERSION F13426.1 GI:710043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 310)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
Lorente, F., Mitchell, H., Marjale, R., Pletu, G., Pouliot, Y.,
Sebastiani, Kabakchis, C. and Tessier, A.

TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE
Tel: 33160778698
Fax: 33160778698
Email: genethon@genethon.fr
Single read.
Genethon library_id: C; Genethon_sequence_id: y2c-2xa02
Insert length: 1500 Std Error: 0.00
Seq primer: (-21)M13 universal
High quality sequence stop: 335.
Location/Qualifiers

FEATURES

source
1. 310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="C-2xa02"
/clone_lib="normalized infant brain cDNA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site-1: HindIII;
Site-2: NotI; sex:Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total
brain; total mRNA was Oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B. Soares, Psychiatry
Dept. Columbia University, USA. Normalization method:
Bento Soares, P.N.A.S. In press"
BASE COUNT 79 a 73 c 89 g 56 t 3 others
ORIGIN

Query Match 4.1% Score 265; DB 21; Length 310;
Best Local Similarity 95.4% Pred. No. 2.8e-40;
Matches 271; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 5115 ggcctcgcggaagaagacgagctgaagagcgtgcagatgcaagagcttgctgaaatgct 5174
27 GGAAGCGCGCGGAGAGAGCGTTAAAGAGCTGCCAGATGCAAGAGCTTGAATATCT 86
Oy 5175 catgagtaagaataaccgacacacacacacacacacacacacacacacacacacac 5234
87 CATGAGTAAGATATACCGCGACTT CAGTACGTTAACCAAGCAACAGACAGACCTTGCAT 146
Oy 5235 gcaatcgacatccccagcgtcagcttggattgaagagagagacagatgagacat 5294
147 GCACATGGCATACCCAGGCTCAGCTTTGATTGAAGAGAGACACCATGAGACAT 206
Oy 5295 tgaatcgacgtccagcgttctgttgaagagatgcttgagagagcagcgaagcaga 5354
207 TGATGATGCTCCAGCGGTCTNTTGAAGAGATGCTGACTTTGGAGGGGCGAGCTAACGAA 266
Oy 5355 gtccctgacggtcactcaacagcgttcaatcaaatcggtcttc 5398
Db 267 GTCCCTGACGGTCACTCACAGCGGTCCATTACCAATGCGTCTC 310

RESULT 12
AM048531/c 365 bp mRNA EST 18-SEP-1999
LOCUS UI-M-BH1-ax-b-07-0-UI.s1 NIH_BMAP_M_S2 Mus musculus cDNA clone
DEFINITION UI-M-BH1-ax-b-07-0-UI 3', mRNA sequence.
ACCESSION AM048531
VERSION AM048531.1 GI:5909060
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S. in press"
BASE COUNT 58 a 50 c 53 g 83 t 8 others
ORIGIN

Query Match 3.8%; Score 242.4; D9 21: Length 252;
Best Local Similarity 96.4%; Pred. No. 4.9e-36;
Matches 243: Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Oy 5162 ttggccacattacc:caagcgggagagagagcctggagacagcctaataatgagcagc 6221
|||||
Db 252 GTGGCCACATTACTCTAGCGGAGGAGCAGAGCTGGACAGCGCAGCTTAAATGAGCAGC 193
6222 tcccaagaatcaactgcctcgaactcccgaggccattgaaagaaacaatccttacgcaaa 6281
|||||
Db 192 TCCCNAGAAATCAGCTGCTCGACTCCCGGGGCCATTGAAAGAAACAATCCTTACGGAAAA 133
Oy 6282 tcttaacacctggtataaacaagacagcatgactggacagcggttgaatataatcaatcaaac 6341
|||||
Db 132 TCTTACACCCCTGGTATNACAGACAGCATGACTGACAGNGGTTGTAATPACAAATCAAAC 73
Oy 6342 aatcaatcaagctaaccttttttaacggaattccaatattataatataaagaatg 6401
|||||
Db 72 AANTCAATNNAAGCTACTTTTTTTACGGAATTCATATTTATATTAAGNAAATTG 13
Oy 6402 ccaaatatataat 6413
|||||
Db 12 CCNNAATATATATT 1

Search completed: April 25, 2000, 16:59:22
Job time: 7998 sec



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm.
Run on: Tue Apr 25 12:27:35 2000: MasPar time 65.06 Seconds
695.328 Million cell updates/sec
Tabular output not generated.

Title: >US-08-956-991-2
Description: (1-1910) from US08956991A.pep
Perfect Score: 13516
Sequence: 1 MHITALLSLFQSFANVFSEDL.....KAIGQVSYICLHLEWTFPC 1910

Scoring table:
PAM 150
Gap 11

Searched: 188963 segs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
l-geneseqp

Statistics: Mean 42.183: Variance 238.791: scale 0.177

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	13516	100.0	1910	1	M42086	Human Down syndrome-ce	0.00e+00
2	11110	87.2	1571	1	M42087	Human Down syndrome-ce	0.00e+00
3	2081	15.4	465	1	M55045	Neutral adhesion molecu	8.93e-142
4	740	5.5	1447	1	R68553	Deleted in colorectal	8.88e-42
5	740	5.5	1447	1	R68553	Deleted in colorectal	8.88e-42
6	629	4.7	1018	1	R87028	Human contactin.	7.39e-34
7	629	4.7	1018	1	R63759	Human contactin (EMBL	1.03e-33
8	623	4.6	1192	1	M57900	Protein of clone C0722	2.80e-33
9	620	4.5	1257	1	M06485	Human LI cell adhesion	4.61e-33
10	606	4.5	1028	1	W28667	Rat contactin ligand f	4.72e-32
11	606	4.5	1028	1	W28667	Rat contactin ligand f	3.39e-32
12	590	4.4	761	1	R92255	Homo sapiens DL185.1 c	6.72e-31
13	588	4.4	1304	1	M59994	Human cell adhesion m	9.35e-31
14	569	4.2	1911	1	M94027	Human neutral cell adhe	2.18e-29
15	569	4.2	1911	1	M27225	Human protein tyrosine	2.18e-29
16	569	4.2	1911	1	R71726	Human PTP-OB.	2.18e-29
17	549	4.1	582	1	R92256	Human cell adhesion m	5.93e-28
18	525	3.9	1070	1	M06747	Human colon carcinoma	3.09e-26
19	514	3.8	753	1	M83927	Human T85 protein.	1.89e-25
20	486	3.6	1501	1	R72858	Rat receptor type-prot	1.86e-23
21	414	3.1	1496	1	M81030	Melanin associated an	2.24e-18
22	384	2.9	400	1	R75203	Tyrosine phosphatase M	5.59e-17
23	353	2.6	1291	1	R75201	Tyrosine phosphatase M	3.89e-14

24	321	2.4	467	1	R84094	Nsk2 receptor with put	6.07e-12
25	319	2.4	467	1	M62575	Alternatively spliced	8.31e-12
26	319	2.4	475	1	R94982	Nsk2 extracellular dom	8.31e-12
27	322	2.4	478	1	R92718	Mouse muscle-localized	5.19e-12
28	322	2.4	478	1	R92718	Mouse muscle-localized	5.19e-12
29	319	2.4	466	1	M62569	Alternatively spliced	8.31e-12
30	319	2.4	466	1	M62569	Alternatively spliced	8.31e-12
31	322	2.4	467	1	M62583	Mouse receptor tyrosin	5.19e-12
32	320	2.4	468	1	R92717	Mouse muscle-localised	7.10e-12
33	320	2.4	468	1	M6507	Rat Dmk receptor.	7.10e-12
34	326	2.4	468	1	M6507	Rat muscle-specific k1	2.77e-12
35	326	2.4	469	1	M6506	Human Dmk receptor	2.77e-12
36	319	2.4	471	1	M26611	Human muscle-specific	8.31e-12
37	319	2.4	471	1	M26611	Nsk2 receptor.	8.31e-12
38	319	2.4	471	1	M62568	Mouse receptor tyrosin	8.31e-12
39	319	2.4	473	1	R84092	Nsk2 receptor with alt	8.31e-12
40	319	2.4	473	1	M62573	Alternatively spliced	8.31e-12
41	319	2.4	481	1	R84091	Nsk2 receptor with alt	8.31e-12
42	320	2.4	481	1	M62572	Mouse Nsk2 (alternativ	8.31e-12
43	330	2.4	1225	1	M52289	Homo sapiens cdo tumou	7.10e-12
44	317	2.3	863	1	M52287	Rattus norvegicus cdo	1.48e-11
45	302	2.2	1091	1	R84088	Nsk2 receptor with alt	1.14e-11
					M41641	Sequence used in detec	1.18e-10

ALIGNMENTS

RESULT	ID	Location/Qualifiers
1	M42086	standard; Protein; 1910 AA.
AC	28-SEP-1998	(first entry)
DE	Human Down syndrome-cell adhesion molecule DS-CAM1.	
KW	DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;	
KW	signal transduction; trisomy 21; mental retardation;	
KW	holoprosencephaly; corpus callosum agenesis;	
KW	schizencephaly; diagnosis; assay; human.	
OS	Homo sapiens.	
FS	Homo sapiens.	
FT	Key	
FT	Peptide	1..23
FT	Protein	/label= Sig_peptide
FT	Protein	24..1910
FT	Protein	/label= Mat_protein
FT	Domain	24..887
FT	Domain	/label= IG
FT	Domain	/note= "Immunoglobulin type-C2 domain"
FT	Domain	888..1594
FT	Domain	/label= FcN
FT	Domain	/note= "Igf1 receptor type III domain"
FT	Domain	1595..1616
FT	Domain	/label= Transmembrane
FT	Domain	1617..1910
FT	Domain	/label= Cytoplasmic
FT	Region	24..126
FT	Region	/label= Ig1
FT	Region	127..225
FT	Region	/label= Ig2
FT	Region	226..316
FT	Region	/label= Ig3
FT	Region	317..409
FT	Region	/label= Ig4
FT	Region	410..506
FT	Region	/label= Ig5
FT	Region	507..603
FT	Region	/label= Ig6
FT	Region	604..697
FT	Region	/label= Ig7
FT	Region	698..792
FT	Region	/label= Ig8
FT	Region	793..887
FT	Region	/label= Ig9
FT	Region	46..102
FT	Disulfide_bond	145..197
FT	Disulfide_bond	246..293
FT	Disulfide_bond	335..385

FT Disulfide_bond 428..484
FT Disulfide_bond 575..575
FT Disulfide_bond 617..669
FT Disulfide_bond 711..766
FT Disulfide_bond 809..865
FT Disulfide_bond 1307..1359
FT Modified_site 78..80
FT /note="Asn is N-glycosylated"
FT Modified_site 106..108
FT /note="Asn is N-glycosylated"
FT Modified_site 470..472
FT /note="Asn is N-glycosylated"
FT Modified_site 487..489
FT /note="Asn is N-glycosylated"
FT Modified_site 658..660
FT /note="Asn is N-glycosylated"
FT Modified_site 666..668
FT /note="Asn is N-glycosylated"
FT Modified_site 710..712
FT /note="Asn is N-glycosylated"
FT Modified_site 748..750
FT /note="Asn is N-glycosylated"
FT Modified_site 795..797
FT /note="Asn is N-glycosylated"
FT Modified_site 924..926
FT /note="Asn is N-glycosylated"
FT Modified_site 1142..1144
FT /note="Asn is N-glycosylated"
FT Modified_site 1160..1162
FT /note="Asn is N-glycosylated"
FT Modified_site 1250..1252
FT /note="Asn is N-glycosylated"
FT Modified_site 1271..1273
FT /note="Asn is N-glycosylated"
FT Modified_site 1324..1326
FT /note="Asn is N-glycosylated"
FT Modified_site 1341..1343
FT /note="Asn is N-glycosylated"
FT Modified_site 1488..1490
FT /note="Asn is N-glycosylated"
FT Modified_site 1488..1490
FT /note="Asn is N-glycosylated"
PN MO9817795-A1.
PN 30-APR-1998.
PR 23-OCT-1997: U19547.
PR 25-OCT-1996: US-029322.
PA (CEDA-) CEDARS STRAI MEDICAL CENT.
PI Korenberg JR:
PI WP1: 98-271791/24.
DR N-PSDB: V31981.
PT New isolated Down's Syndrome-cell adhesion molecule - used to
PT develop products for detection, diagnosis and therapy of
PT developmental and neurological abnormalities
PS Claim 2: Page 73-78: 109pp: English.
CC This polypeptide comprises Down syndrome-cell adhesion molecule
CC DS-CAM1, a cell surface glycoprotein belonging to a novel subclass
CC of the Ig superfamily with highest homology to neural cell adhesion
CC molecules. Its amino acid sequence was deduced from cDNA clones
CC (see V31981) isolated from a trisomy 21 foetal brain library. A
CC splice variant DS-CAM2 (see M42087), which is non-membrane bound
CC was also identified. The invention also provides human and murine
CC DS-CAM nucleic acid sequences (see also V31981-88), expression
CC vectors and host cells, transgenic animals, antibodies, antisense
CC oligonucleotides, and primers derived from DS-CAM nucleic acid.
CC DS-CAM polypeptides are associated with developmental and
CC neurological processes. They can be used in e.g. neural prosthetic
CC devices used in entubulation methods of repairing (regenerating)
CC damaged or severed peripheral nerves, and also in bioassays to
CC identify agonists and antagonists. The products can also be
CC used in detection, diagnosis and therapy of developmental and
CC neurological abnormalities such as Down syndrome, mental
CC retardation, holoprosencephaly, agenesis of the corpus callosum,
CC or schizencephaly.
CC Sequence 1910 AA:

Query Match 100.0%: Score 13516; DB 1; Length 1910;
Best Local Similarity 100.0%: Pred. No. 0.00e+00;
Matches 1910: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 MWILALSLFQSFANFSEDLHSLYFNASLQEVYFASTGTGLVPCAAACIPVTLRWYL 60
1 MWILALSLFQSFANFSEDLHSLYFNASLQEVYFASTGTGLVPCAAACIPVTLRWYL 60
Db 61 ATGEETDYDVGIRHVNPGTGLQIFPPSPSSTLIHNTTYCTAENPSGKIRSDVHIKA 120
61 ATGEETDYDVGIRHVNPGTGLQIFPPSPSSTLIHNTTYCTAENPSGKIRSDVHIKA 120
Qy 61 ATGEETDYDVGIRHVNPGTGLQIFPPSPSSTLIHNTTYCTAENPSGKIRSDVHIKA 120
121 VIREPTVREDDQKTRGNVAVFKCIIPSSVEAYITVSWEKDTVLSYSGSRFLITSGA 180
121 VIREPTVREDDQKTRGNVAVFKCIIPSSVEAYITVSWEKDTVLSYSGSRFLITSGA 180
Db 181 LYIKDVQNEDEGLYNYRCITRRHRYTGETROSNSARLFVSDPNASPSILDGFRKMAAGQ 240
181 LYIKDVQNEDEGLYNYRCITRRHRYTGETROSNSARLFVSDPNASPSILDGFRKMAAGQ 240
Qy 181 LYIKDVQNEDEGLYNYRCITRRHRYTGETROSNSARLFVSDPNASPSILDGFRKMAAGQ 240
241 RVELPCALGHPEDDYRWLKDNNPLELSGRFQKVTGLLIENIRPSDSGSYCEVSNRYG 300
241 RVELPCALGHPEDDYRWLKDNNPLELSGRFQKVTGLLIENIRPSDSGSYCEVSNRYG 300
Db 301 TAKYIGRLYKQPLKATISPRKYKSSVGSVSLSSVGTEDQELSWYRNGELINPKNV 360
301 TAKYIGRLYKQPLKATISPRKYKSSVGSVSLSSVGTEDQELSWYRNGELINPKNV 360
Qy 301 TAKYIGRLYKQPLKATISPRKYKSSVGSVSLSSVGTEDQELSWYRNGELINPKNV 360
361 RITGINHENLIMHMYSDGAYQCFVRKDKLSADQYVQVLEDDPKIISAFSEKRVSP 420
361 RITGINHENLIMHMYSDGAYQCFVRKDKLSADQYVQVLEDDPKIISAFSEKRVSP 420
Db 421 AEPVSLMCNKVGTPLTITWTLUDDPLKGGSHRISOMLTSGNVSYNINISSGVRODG 480
421 AEPVSLMCNKVGTPLTITWTLUDDPLKGGSHRISOMLTSGNVSYNINISSGVRODG 480
Qy 421 AEPVSLMCNKVGTPLTITWTLUDDPLKGGSHRISOMLTSGNVSYNINISSGVRODG 480
481 VYRCANNSAGVYVYQARINVRGPASIRPKNITAIAGRDYTHCAVIGPYYSIKWYN 540
481 VYRCANNSAGVYVYQARINVRGPASIRPKNITAIAGRDYTHCAVIGPYYSIKWYN 540
Db 541 SNLDPFNHROVAFENNGTKLSVDQKEVDEGTYCNLVQPOLSTOSVHYVTKVPEFIQ 600
541 SNLDPFNHROVAFENNGTKLSVDQKEVDEGTYCNLVQPOLSTOSVHYVTKVPEFIQ 600
Qy 541 SNLDPFNHROVAFENNGTKLSVDQKEVDEGTYCNLVQPOLSTOSVHYVTKVPEFIQ 600
601 PREPPRFSGQVFIQCVVVSGLDPLITTWQKDRPIPGSLGVTIDNIDFTSLRISNLS 660
601 PREPPRFSGQVFIQCVVVSGLDPLITTWQKDRPIPGSLGVTIDNIDFTSLRISNLS 660
Db 661 LMHNGNYTCIARNEAAVEHQSOLIVRVPKFEVQPRDODIYKAVILNCSAEGYVPPT 720
661 LMHNGNYTCIARNEAAVEHQSOLIVRVPKFEVQPRDODIYKAVILNCSAEGYVPPT 720
Qy 661 LMHNGNYTCIARNEAAVEHQSOLIVRVPKFEVQPRDODIYKAVILNCSAEGYVPPT 720
721 IYVKEFKAGVYQFOPIALNGRIQVLSNGSLIKHYVEDSGYLCKVSDVADYSKSM 780
721 IYVKEFKAGVYQFOPIALNGRIQVLSNGSLIKHYVEDSGYLCKVSDVADYSKSM 780
Qy 721 IYVKEFKAGVYQFOPIALNGRIQVLSNGSLIKHYVEDSGYLCKVSDVADYSKSM 780
781 YLVKVPKPAITSPNTTLATQCKEMKCTAHEKPIIVREKEDRIINPEMAYLYSTK 840
781 YLVKVPKPAITSPNTTLATQCKEMKCTAHEKPIIVREKEDRIINPEMAYLYSTK 840
Db 841 EYGEVISTQLIIPYREDSGFSCHAINSYGEDGIIQIVQEPDPELEINDVART 900
841 EYGEVISTQLIIPYREDSGFSCHAINSYGEDGIIQIVQEPDPELEINDVART 900
Qy 841 EYGEVISTQLIIPYREDSGFSCHAINSYGEDGIIQIVQEPDPELEINDVART 900
901 IITLRMTGFGNGSPITGYDIECKNKSQSDMSAQRKDVSPQNSATIIDIHPSSTYSIRM 960
901 IITLRMTGFGNGSPITGYDIECKNKSQSDMSAQRKDVSPQNSATIIDIHPSSTYSIRM 960
Db 961 YAKNRICKSEPSNELITADDEAAPDQPEVHLPEIPSSQSIIVTKAKPKHLQNGIINGY 1020
961 YAKNRICKSEPSNELITADDEAAPDQPEVHLPEIPSSQSIIVTKAKPKHLQNGIINGY 1020
Qy 961 YAKNRICKSEPSNELITADDEAAPDQPEVHLPEIPSSQSIIVTKAKPKHLQNGIINGY 1020
1021 QIGYREVSTGNGQFNITISVDTSGDSEVYTLDMLNKFTQYGLVVOACNRAGTGSSOELI 1080

 M P E E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Apr 25 13:02:51 2000; Maspar time 28.68 Seconds
 862,987 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-956-991-2
 Description: (1-1910) from US08956991A.pep
 Perfect Score: 13516
 Sequence: 1 MMTLALSLTQSFANFSEDL.....KAIGVTSYICLTLEWTF 1910

Scoring table:
 PAM 150
 Gap 11

Searched: 131253 seqs, 12956647 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-issued
 1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 39.388; Variance 240.403; scale 0.164
 Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2081	15.4	465	2	US-08-752-	Sequence 5, Applicatio	2,74e-132
2	2071	15.3	462	2	US-08-752-	Sequence 7, Applicatio	1,37e-131
3	740	5.5	1447	3	PCT-US94-0	Sequence 2, Applicatio	2,64e-19
4	629	4.7	1018	1	US-08-408-	Sequence 6, Applicatio	8,67e-32
5	629	4.7	1018	1	US-08-408-	Sequence 6, Applicatio	8,67e-32
6	629	4.7	1018	1	US-08-714-	Sequence 6, Applicatio	8,67e-32
7	606	4.5	1018	1	US-08-452-	Sequence 2, Applicatio	3,07e-30
8	569	4.2	1911	1	US-08-348-	Sequence 5, Applicatio	9,38e-28
9	569	4.2	1911	2	US-08-800-	Sequence 5, Applicatio	9,38e-28
10	569	4.2	1911	2	PCT-US94-1	Sequence 5, Applicatio	9,38e-28
11	486	3.6	1501	2	US-08-716-	Sequence 3, Applicatio	3,23e-22
12	486	3.6	1501	2	US-08-447-	Sequence 3, Applicatio	3,23e-22
13	449	3.3	596	2	US-08-752-	Sequence 13, Applicati	9,06e-16
14	448	3.3	612	2	US-08-752-	Sequence 11, Applicati	1,05e-19
15	448	3.3	612	2	US-08-752-	Sequence 12, Applicati	5,78e-16
16	381	2.8	605	2	US-08-752-	Sequence 8, Applicatio	2,59e-15
17	379	2.8	946	3	PCT-US95-0	Sequence 13, Applicati	3,48e-15
18	361	2.7	946	2	US-08-752-	Sequence 9, Applicatio	5,13e-14
19	359	2.7	630	2	US-08-752-	Sequence 14, Applicati	6,92e-14
20	322	2.4	478	3	PCT-US95-0	Sequence 15, Applicati	1,66e-11
21	322	2.4	860	3	PCT-US95-0	Sequence 19, Applicati	1,66e-11
22	322	2.4	868	3	PCT-US95-0	Sequence 21, Applicati	1,66e-11
23	320	2.4	868	1	US-08-374-	Sequence 1, Applicatio	2,23e-11

24	320	2.4	868	2	US-08-644-	Sequence 1, Applicatio	2,23e-11
25	326	2.4	869	2	US-08-374-	Sequence 16, Applicati	9,20e-12
26	326	2.4	869	1	US-08-644-	Sequence 29, Applicati	9,20e-12
27	282	2.1	2231	1	US-08-153-	Sequence 16, Applicati	5,74e-09
28	262	1.9	2324	1	US-08-283-	Sequence 1, Applicatio	1,03e-07
29	262	1.9	2324	3	PCT-US95-0	Sequence 1, Applicatio	1,03e-07
30	262	1.9	2327	4	5455158-1	Patent No. 5455158.	1,03e-07
31	262	1.9	2386	2	US-09-016-	Sequence 12, Applicati	1,03e-07
32	262	1.9	2446	2	US-08-551-	Sequence 2, Applicatio	1,03e-07
33	262	1.9	2446	3	PCT-US93-1	Sequence 2, Applicatio	1,03e-07
34	233	1.7	252	2	US-08-414-	Sequence 57, Applicati	6,38e-06
35	233	1.7	287	2	US-08-414-	Sequence 49, Applicati	4,81e-06
36	229	1.7	287	2	US-08-414-	Sequence 48, Applicati	1,12e-05
37	229	1.7	304	2	US-08-414-	Sequence 44, Applicati	1,12e-05
38	229	1.7	310	2	US-08-414-	Sequence 45, Applicati	4,81e-06
39	235	1.7	315	2	US-08-414-	Sequence 47, Applicati	4,81e-06
40	235	1.7	338	2	US-08-414-	Sequence 43, Applicati	4,81e-06
41	235	1.7	338	2	US-08-414-	Sequence 42, Applicati	4,81e-06
42	234	1.7	419	4	5169835-2	Patent No. 5169835.	5,54e-06
43	232	1.7	549	2	US-08-836-	Sequence 11, Applicati	7,35e-06
44	232	1.7	574	2	US-08-836-	Sequence 21, Applicati	7,35e-06
45	234	1.7	828	1	US-08-261-	Sequence 2, Applicatio	5,54e-06

ALIGNMENTS

RESULT 1
 ID US-08-752-307B-5 STANDARD: PRT: 465 AA.
 AC xxxxxx
 DE Sequence 5 Application US/08752307B
 XX Patent No. 5952171
 CC GENERAL INFORMATION:
 CC APPLICANT: McCarthy, Sean A.
 CC APPLICANT: Geating, David P.
 CC APPLICANT: Levinson, Douglas A.
 CC TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
 CC NUMBER OF SEQUENCES: 14
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESS: Fish & Richardson, P.C.
 CC STREET: 225 Franklin Street
 CC CITY: Boston
 CC STATE: MA
 CC COUNTRY: US
 CC ZIP: 02110-2804
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: diskette
 CC OPERATING SYSTEM: IBM Compatible
 CC SOFTWARE: FASTSEQ for Windows Version 2.0
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/752,307B
 CC FILING DATE: 19-NOV-1996
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER:
 CC FILING DATE:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Melkijohn, Ph.D., Anita L.
 CC REGISTRATION NUMBER: 35,283
 CC REFERENCE/DOCKET NUMBER: 09404/020001
 CC TELEPHONE: 617-542-5070
 CC TELEFAX: 617-542-8906
 CC TELEX: 200154
 CC INFORMATION FOR SEQ ID NO: 5:
 CC SEQUENCE CHARACTERISTICS:

[illegible]

```

CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/752,307B
CC FILING DATE: 19-NOV-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MeljoJohn, Ph.D., Anita L.
CC REGISTRATION NUMBER: 35,283
CC REFERENCE/DOCKET NUMBER: 09404/020001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-542-5070
CC TELEFAX: 617-542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ. ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 462 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE 462 AA: 50937 MW: 1144658 CN:

Query Match 15.3%; Score 2071; DB 2; Length 462;
Best Local Similarity 57.4%; Pred. No. 1,37e-131;
Matches 263; Conservative 104; Mismatches 90; Indels 1; Gaps 1.

Cc 1 MWLVLTLLDLSLHKAAPEDGTSILYFVNLSLOOVFSSSGVGVVPCPAGSPALRLMY 60
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 1 MWLLA-LSLFSQFANVFSEDLHSLSTFYNASLDEYVFASTGTLVPCPAGIPPLYLRLMY 59
Cc
Cc 61 LATGDIDYDVPHIRHVHANGTLQDLPFSPSAFNSFIHDNDFCTAENAGKIRSFNIRVK 120
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 60 LATGEELIYVPGIRHVHNGTLQIFPPSPSSFLLIHDNYYCTAENPSGKIRSDVHIK 119
Cc
Cc 121 AVREPEYTVRVEDQSGMGNVAFKCLIPSSVOEYVYSKMDTVSIIIPENRFITYHG 180
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 120 AVLEPEYTVRVEDQSGMGNVAFKCLIPSSVEAYITVWSKMDTVSLVSGSRFLITSTG 179
Cc
Cc 181 GLYISDVQKEDALSTYRCITKHKYSGETROSGCARLSVTDPAESIPTILDFGRHSEVWAG 240
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 180 ALVIKDVQKEDGLYVNRCTIRRTARTGETROSNARLFLVSDPANSAPSLTDFGRHKAMAG 239
Cc
Cc 241 HTYVELPCTASGPIPAIRMLKDGRLPADSRMWRKRTIGLTISDLRTEDSGTICYVTNF 300
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 240 QRELPEKALGHPEDPEYRWIKMNPMLDELSGRFQKTYVTGLIENIRPSDGSVYCEVSRY 299
Cc
Cc 301 GSMFANGILMWIDVLPVLTLPYPRKLKTIGSTVILSCALTGSPFTIRWRNTELLPDEA 360
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 300 GTAKYVIGRLYKOPLRKATISPKRYKSSVSGVSLSCSVTETDEQELSWRNEILLPGKN 359
Cc
Cc 361 ISIRGSLNETLLTTSAKSHSGAYOCFAETRRAGTADQFAIALLEDGTPIRVSSSEKVVN 420
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 360 VRTIGIHNEMLIMDHWKSDGAYOCFPRKDKLSAODYVGVVLEDTPKIISAFSEKVV 419
Cc
Cc 421 PGDEFSLMCAAGAPRPYTYMALDDEPIRVGRSHRTNQ 458
Cc |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cc 420 PAEPVSLMCNWKGTPLPTITWLTDDDPILKGGSHRISQ 457
Cc
Cc
Cc RESULT 3 STANDARD: PRT: 1447 AA.
Cc ID PCR-US94-05277-2
Cc XX xxxxxx
Cc DT
Cc XX
Cc DE Sequence 2, Application PC/TUS9405277
Cc CC Sequence 2, Application PC/TUS9405277
Cc CC GENERAL INFORMATION:
Cc CC APPLICANT: Bruskin, Arthur
Cc CC

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 14:46:32 : Search time 7810.03 Seconds
(without alignments)
-2567.504 Million cell updates/sec

Title: US-08-956-991-1

Perfect score: 6604
1 tgcactgagcgcgagcagcgcg.....gaaattgcccaaatattt 6604

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 82193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vl:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_of:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vl:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*

45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6212	94.1	6413	11	AF023450	AF023450 Homo sapi
2	6110	92.5	6110	11	AF023449	AF023449 Homo sapi
3	2051.2	31.1	6025	10	AB032858	AB032858 Homo sapi
4	699	10.6	721	11	H0MY291F03	AF086100 Homo sapi
5	495.4	7.5	109866	11	AF043945	AF043945 Homo sapi
6	469.4	7.1	145861	11	AF064862	AF064862 Homo sapi
7	358.8	5.4	430	13	G36681	G36681 SHGC-53823
8	319.4	4.8	119678	11	AF064866	AF064866 Homo sapi
9	319.4	4.8	120007	11	AF064864	AF064864 Homo sapi
10	307.2	4.7	121019	11	AF042091	AF042091 Homo sapi
11	307.2	4.7	155407	11	AF042090	AF042090 Homo sapi
12	296	4.5	159424	11	AF064865	AF064865 Homo sapi
13	154.6	2.3	189607	32	AP000757	AP000757 Homo sapi
14	154.6	2.3	223724	32	AP000834	AP000834 Homo sapi
15	153.6	2.3	193789	32	AP000849	AP000849 Homo sapi
16	152.8	2.3	124395	44	AC013397	AC013397 Homo sapi
17	121.4	1.8	193789	32	AP000849	AP000849 Homo sapi
18	98	1.5	96870	41	AC009739	AC009739 Drosophila
19	96.4	1.5	279137	41	AC008540	AC008540 Homo sapi
20	91.8	1.4	110000	45	AC017014_0	AC017014 Homo sapi
21	91.2	1.4	174707	45	AC017082	AC017082 Homo sapi
22	89.2	1.4	107475	45	AC017106	AC017106 Homo sapi
23	89	1.3	7606	34	DROLARM	AC017106 Homo sapi
24	89	1.3	31646	45	AC017967	M27700 D.melanogaster
25	89	1.3	90563	42	AC010115	AC017967 Drosophila
26	88.4	1.3	110000	45	AC017014_1	AC010115 Drosophila
27	87.4	1.3	20940	43	AC014917	Continuation (2 of AC014917 Drosophila)
28	87.2	1.3	87042	43	AC011885	AC014917 Drosophila
29	86.6	1.3	110000	45	AC017014_1	AC011885 Homo sapi
30	86.6	1.3	141025	44	AC015693	Continuation (2 of AC015693 Homo sapi)
31	86.4	1.3	181634	44	AC017015	AC015693 Homo sapi
32	86	1.3	96106	45	AC017015	AC017015 Homo sapi
33	83.4	1.3	162548	43	AC010827	AC017015 Homo sapi
34	83.2	1.3	147343	44	AC011871	AC010827 Homo sapi
35	83.2	1.3	173721	43	AC015866	AC011871 Homo sapi
36	83	1.3	107475	45	AC017106	AC015866 Homo sapi
37	82.8	1.3	840	8	CNS01BNS	AC017106 Homo sapi
38	81.4	1.2	56416	41	AC011152	AL114464 Botrytis
39	80.4	1.2	182110	44	AC005302	AC011152 Homo sapi
40	80.4	1.2	194355	33	AL133216	AC005302 Mus muscu
41	80.2	1.2	139207	41	AC011429	AL133216 Homo sapi
42	80.2	1.2	159836	44	AC008334	AC011429 Homo sapi
43	79.6	1.2	181634	44	AC015623	AC008334 Drosophila
44	79.6	1.2	60539	44	AC015463	AC015623 Homo sapi
45	79.4	1.2	67220	45	AC017014_3	AC015463 Homo sapi

ALIGNMENTS

RESULT 1
AF023450 AF023450 6413 bp mRNA
LOCUS Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (DSCAM)
DEFINITION mRNA, complete cds.
ACCESSION AF023450
VERSION AF023450.1 GI:3169767

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 6413)
Yamakawa,K., Huo,Y.-K., Haendel,M.A., Hubert,R., Chen,X.-N.,
Lyons,G.E. and Korenberg,J.R.
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a
Down Syndrome Region and is Involved in the Development of the
Nervous System
Unpublished
2 (bases 1 to 6413)
Yamakawa,K., Huo,Y.-K., Haendel,M.A., Hubert,R., Chen,X.-N.,
Lyons,G.E. and Korenberg,J.R.
Direct Submission
Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los
Angeles, CA 90048-1869, USA
Location/Qualifiers
1..6413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21 (trisomy 21)"
/map="21q22, between HMG14 and MX1"
/tissue="brain"
/dev_stage="14 weeks, fetal"
/clone="CHD2-52"
/note="derived from alternately-spliced mRNA"
1..6413
/gene="DSCAM"
453..5168
/note="member of immunoglobulin superfamily; involved in
nervous system development"
/product="Down syndrome cell adhesion molecule"
/protein_id="AAC17967.1"
/db_xref="GI:3169768"
/translation="MW141SLSPANFSEDLHSSLYFNVASLOEVVFASTGTYV
PCPAGIPPPVTLRWYLATGEIYDPGIRHVPNTLOIFPPSPSEFTLIHNTTYC
TENSGKIRSDOVHAKAVREPVRVEDOKTMGNAVEFCIIIPSSVEAIVTWSM
EDVTSVSGSRFLITSTGALIKDVQNDGLYNRCTIRHYTETEQSARLTVS
DPANAPSLIDGFDHRKAMAGORVELPCALGPDPDWMLNDLPSRFOVTV
GLLIENIRPSDGSYCEVSNRYGTAKVRLYVOPKATISPKVSSVSGVOSLS
CSVTGEDEL.SWYRMGELINPGKNVRIIGNHMLINDHMYSGGAGVGFVRDKL
SNADVOYVLEDGTPIKIIISAEKVSPAEPVSLMCNKIKPTLPITTTLLDDPLK
GSHRISOMTISEGNVSTYLNISSOVRODGVTRCTANNSAGVLYOARIINRGPASIR
PKMNTAIAGRDYLHCRVIGPYYSIKYKNSNLPFNHRQVARENGTLKLSVOK
EYDEGEYTCNVLPOLSTOSVHTVVPPIQEPFEPREFICQRFVIPCVAVSGDL
PLITWOKDGRPIGSLGVTIDNIDFTSLRISNLSLHNGNYTICIANEAAVYHOS
GLIVRAPKFEVVPQDQDYGKAVLINCAGVAPVITWKFSGACVPOFPAIN
GLIOVLSNGLIKHVVREDSGVYLCKYNDGAVSKMLYTVYIPAMITSYPTTL
ANOGOKREKSTAHGEKPIIVWEKEDRIINEMARIYSTRKEVEEYISTLIQILPY
RDSGSGFSCHAINSDEGRILQLTVQEPDPELEIDVAKRTILMTWGFQNSP
ITGYDECKNSDMSDORVDVPOLSAIIIDHPSSTYSIMVAKNIGKSEPS
NELITITADEAADGDPQEVHLEPIISSQIRVTKAPKRLHGLIIRGOIYREYSDG
GNFOENIISVDTSQDEEYTLDNLNKFTOYGVQACRAGGSPSSOILITTELDVP
SYPERVOAITSPEISISWSTLSKELNGLQGFVITYANAMDMGSLGKIKNTTT
OSLELDLEKTYNSIOVLATFRAGDVSQIOTRKEVDVPPRPAKAAASASV
VYVSWPLPLKNGIIRKTIIVFCSHPPYVVISFEFSPSESTRIPNLNRNQQVSWV
AVTSAGKNSSEIITVEPLAKAPARILFTSGVTTPMKMDIVLPCKAVAGDSPAOKM
KOSNGPSLVNITIDGRSIFSNSTIRIVTKAEDSYSCIANNNMGSEIILNLOVY
PDPORLTVSKTSSITLSLMPGNDGSGSYGLIYSEDSQMGSEFPIPSERSY
RLENLKGTWYKFTLLAONGVGRGRISEIETKATIGKSDPSKDELPASTINTVRIL
NIGNDDGCPITFTLEIYRPGCTTWTTAQNTSLSKSYLLYDLEATWYELQMRVCN
SHGCAEKQKEARCKERS"

BASE COUNT 1633 a 1779 c 1709 g 1292 t
ORIGIN

Query Match 94.1% Score 6212: DB 11: Length 6413:
Best Local Similarity 97.1% Pred. No. 0:

Matches 6413: Conservative 0: Mismatches 0: Indels 191: Gaps 1:

QY 1 tgactgaagccggaagcaagcaatgaagcttgcgcgcgcctgtctgtatgcg 60
DB 1 TGACTGAGCGCGGAGACAGCGCAAGATGAGCTCCCGCGCTGCTGCTGATCGG 60
QY 61 aggtgaagagctgcgaaggaagccgtgcgcgcgcctgtctgtatgcgcg 120
DB 61 AGGTGAGAGGCTGGGCGACGGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 121 gatgagctcacgc 180
DB 121 GATGAGCTCACGCCCGCGCTGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 181 agctgaagc 240
DB 181 AGCTGAGCGCGCACCTGGCGCTCCGCGCGCGAGCGCGCGCGCGCGCGAGTGT 240
QY 241 gaaggaagc 300
DB 241 GCAGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 300
QY 301 ccggaagcgaagc 360
DB 301 CCGGCGCGCTAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 caggtgc 420
DB 361 CAGGTGCGGAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 421 tcgctcgc 480
DB 421 TCGCTCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 480
QY 481 agagcttcgcgaatgtttcagtgaaagcctacacctcagccttacttgcata 540
DB 481 AGAGCTTCGCAATGTTTTCAGTGAAAGACCTACCTCAGCTCTACTTGTCAATG 540
QY 541 ctgcgaagaaag.agtgttct~agacacacgcgcgcgcgcgcgcgcgcgcgc 600
DB 541 CTGCGAAGAGTGTGTTTTCAGACACAGCGGAGCTGTGTCCTGCTGCGCGAG 600
QY 601 gcatccctccgttactcctcagatgtaacctagcaagcgaggaagatcagatgc 660
DB 601 GCATCCCTCCGTACTCTCAGATGTAAGTACTAGCACAGGCGAGAGATTTACGAT 660
QY 661 ccggaagtcgc 720
DB 661 CCGGAGTCGCGCACGTCGCATCCGACGCACTCTCCAAATTTCCCTTCCCTTCA 720
QY 721 gcttcagacttaatacgaataataataataataataataataataataata 780
DB 721 GCTTCAGACTTAATCAATGATATATATTATTTGACAGCTGAAATCTTCAAGGA 780
QY 781 aaatagaagtcagatgcacacacacacacacacacacacacacacacacac 840
DB 781 AAATTAGAAGTCAGAGATGTCACATCAAGGCTGTTAGGAGACCTATACAGTCC 840
QY 841 tggaggaagcgaagaaacatgaaggaatgtgcgcgtctcgaagtgcattacc 900
DB 841 TGGAGGACCAAAACATGAGAGCAATGTTGGGCTTCAAGTGCATATATCCCTCC 900
QY 901 cggaggaagcgaatataactgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
DB 901 CCGTGAAGCGCTTACATACCTGCTCTCATGAGGAGAAAGACCTGTTTCACTGTCT 960
QY 961 gatctagaattctcatcacatccacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
DB 961 GATCTAGATTCTCATGACATCCACGGGAGCCTTGTATATTAAGATGTACAGATGA 1020
QY 1021 atgattgtataactacgcgtgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
DB 1021 ATGATTGTATACTACCGCTGCATACGCGGAGATGATACCGGAGAGACGAGCAGA 1080

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 15:04:57 ; Search time 383.18 Seconds
(without alignments)
4311.991 million cell updates/sec

Title: US-08-956-991-1

Perfect score: 6604
Sequence: 1 tgactgagccgagcagcag.....gaaatgcacaaatatatt 6604

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 segs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6599.2	99.9	6604	1 V31981	Human Down syndrom
2	6208.8	94.0	6413	1 V31988	Human Down syndrom
3	1141.2	17.3	2173	1 V31987	Mouse Down syndrom
4	670.2	10.1	842	1 V31985	Mouse Down syndrom
5	541	8.2	1493	1 V27207	CDNA clone eth001
6	502	7.6	898	1 V31986	Mouse Down syndrom
7	388	5.9	388	1 V31982	Human Down syndrom
8	79.6	1.2	114955	1 X53491	Human adenosine A1
9	77.4	1.2	270	1 X53491	EST clone DY106, N
10	76	1.1	114955	1 X53491	Human adenosine A1
11	73	1.1	3198	1 X02974	Human IL-1ra BAC c
12	66.4	1.0	6000	1 Q86478	Human protein tyro
13	66.4	1.0	6000	1 T85389	Human protein tyro
14	66.4	1.0	6000	1 X06095	Human protein tyro
15	52.2	0.9	801	1 V73801	KSHV IUR terminal
16	62.2	0.9	6225	1 X55273	Human enzyme-relat
17	62	0.9	4257	1 V10362	Infected cell prot
18	62	0.9	4257	1 V68520	The nucleotide seq
19	62	0.9	12001	1 V62176	HSV L/ST region. H
20	62	0.9	117213	1 V62176	HSV-2 strain S15 C
21	61.6	0.9	2218	1 O47929	Paired basic amino
22	61.6	0.9	4403	1 O47927	Orf virus genomic
23	60.4	0.9	4020	1 T91361	DNA encoding pseud
24	60.4	0.9	8438	1 V73500	HSV-2 strain S15 C
25	60.4	0.9	117213	1 V62176	MUSDUJUNX, a sample
26	60.2	0.9	1026	1 O66091	Infected cell prot
27	59.4	0.9	4257	1 V10362	The nucleotide seq
28	59.4	0.9	4257	1 V68520	HSV L/ST region. H
29	59.4	0.9	12001	1 T78508	Platenolide syntha
30	58.8	0.9	44377	1 T78508	Platenolide syntha
31	58.8	0.9	44377	1 T80414	M. scrofulaceum sh
32	58.4	0.9	2504	1 O36467	Mycobacterium tube
33	58.4	0.9	3946	1 T93610	Brn-3a polynucleot
34	58.2	0.9	1266	1 X09011	

C	35	58.2	0.9	2974	1	079630	Partial FRAKE regl
C	36	57.8	0.9	6413	1	V31988	Human Down syndrom
C	37	57.8	0.9	6604	1	V31981	Human Down syndrom
C	38	57.4	0.9	1743	1	T85537	hNET cDNA sequence
C	39	57.4	0.9	1743	1	V16306	CDNA encoding huma
C	40	57.4	0.9	1974	1	V16346	Nucleotide sequenc
C	41	57.4	0.9	6803	1	T85473	Genomic hNET sequ
C	42	57.4	0.9	6803	1	V16305	Genomic DNA encodi
C	43	57.2	0.9	1548	1	V33277	Nucleotide sequenc
C	44	57.2	0.9	1581	1	V33278	Nucleotide sequenc
C	45	56.8	0.9	2538	1	X06987	Mouse neuregulin r

ALIGNMENTS

RESULT 1
V31981
ID V31981 standard: cDNA; 6604 BP.
AC V31981:
DT 28-SEP-1998 (first entry)
DE Human Down syndrome-cell adhesion molecule DS-CAM1 cDNA.
KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
KW signal transduction; trisomy 21; mental retardation;
KW holoprosencephaly; corpus callosum agenesis;
KW schizencephaly; diagnosis: assay; human; ds; ss.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT CDS
FT /tag= a
FT sig_peptide 452..521
FT /tag= b
FT mat_peptide 522..6162
FT /tag= c
FT
FT W09817795-A1.
PD 30-APR-1998.
PE 23-OCT-1997; U19547.
PR 25-OCT-1996; US-029322.
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PI Korenberg JR;
PI WPI: 98-271791/24.
DR P-PSDB: W42086.
PT New isolated Down's Syndrome-cell adhesion molecule - used to
PT develop products for detection, diagnosis and therapy of
PT developmental and neurological abnormalities
PS Claim 3; Page 65-72; 109pp; English.
CC This cDNA clone codes for Down syndrome-cell adhesion molecule
CC DS-CAM1 (see W42086), a cell surface glycoprotein belonging to a
CC novel subclass of the Ia superfamily with highest homology to
CC neural cell adhesion molecules. A modified direct cDNA selection
CC technique was applied to bacterial and pl artificial chromosomes
CC between EMS2 and MX1 by using cDNA from trisomy 21 human foetal
CC brain. A unique cDNA fragment, designated E51 (see V31982), was
CC obtained and used to screen a trisomy 21 human foetal brain (14-wk
CC gestation) cDNA library. Further clones were obtained by exon
CC trapping. The complete DS-CAM1 cDNA sequence was deduced from
CC overlapping clones. A splice variant cDNA (see V31988), encoding
CC non-membrane bound DS-CAM2 (see W42087), was also identified. The
CC DS-CAM gene spans 900-1200 kb of genomic DNA and maps at chromosome
CC 21q22.2-22.3. The invention also provides murine DS-CAM partial
CC sequences (see V31985-87), expression vectors and host cells,
CC transgenic animals, antisense oligonucleotides, and primers useful
CC for amplification of DS-CAM nucleic acid. DS-CAM polypeptides are
CC associated with developmental and neurological processes. They can
CC be used in e.g. neural prosthetic devices used in entubation
CC methods of repairing (regenerating) damaged or severed peripheral
CC nerves. The products can also be used in detection, diagnosis and
CC therapy of developmental and neurological abnormalities such as
CC Down syndrome, mental retardation, holoprosencephaly, agenesis of
CC the corpus callosum, or schizencephaly. Antisense oligonucleotides
CC are used for inhibiting translation of mRNA. 1768 G; 1336 T;
SQ Sequence 6604 BP; 1673 A; 1827 C; 1768 G; 1336 T;

Query Match	99.9%	Score 6599.2;	DB 1;	Length 6604;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 6601: Conservative	0;	Mismatches	3;	Gaps 0;

[illegible]

QY	1021	atggtattggtataactaacctgcgtctgcatcagcggaatcgcataacacccggagaagacagagcgaga	1080
Db	1021	ATGGATTGTATTACTACACCGCTGCATACCGCGGCATCGATACACCGGAAGACGAGCGACGA	1080
QY	1081	gcaacagcgcgaagactctttgtaatacagaccagcgaaactcagccccatccatactcgtatg	1140
Db	1081	GCACACAGCCGACAGCTTTTGTATACGACCAGCCAGACTCACCCCCATTCATACGTGATG	1140
QY	1141	ggtttgcacatcgcgaagaagccatgagctgggacggtgtgtgagctgtcccttgcgaagcgtcg	1200
Db	1141	GGTTTGCACATCCGAAGAAGCCATGGCTGGGACAGCGTGTGGAGCTGCCCTTGCAGAAGCGCTCG	1200
QY	1201	ggagcccttgagccagaattaccgctgtgctctaaagaagaacatgtccctctgtaactcttcagga	1260
Db	1201	GGCAGCCCTTGAGCCAGATTACCGCTGCCTGADAGACAACTGCCCTTGAACTTTCAGGGA	1260
QY	1261	ggttcacagaagaccgttgaacgggctgtcattcttgagaacatctgcgccctcggactoaagca	1320
Db	1261	GGTTCACAGAAGACCGGTGACGGGGCTGCTCATTTGAGAACATTGGCCCTCGGACTGAGCA	1320
QY	1321	gctatgtcttcttgaaagtgtctccaacagatacggaaactgctaagggtgataaggccgctgtacg	1380
Db	1321	GCTATGTCTTGTGAAGTGTCCAAACAGATACGGAACGTCTAAGGTGATAGCGCCGCTGATCG	1380
QY	1381	tgaaacagccacattgaagaagccaccatcaatcccccagaagaattaaagcagctgtggtaacc	1440
Db	1381	TGAACACGCCACTGAAGAAGCCACCATCACTGCCAGAGAGTTAAAGCAGCGTGGGTACCC	1440
QY	1441	aagttccctgtgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1500
Db	1441	AAGTTCTCTGTCTCTCTGTCGAGGTGTCACGAGAACTGAGACCGAGAACCTCTCTGTGACCCGA	1500
QY	1501	atgtgtgaatccctcccaacccctcgggaaaaaatgtaagatacacaagagatacacaacgaaaaaac	1560
Db	1501	ATGTTGAATCTCTCAACCTCGGAATAAAATGTGAGATCTCAGAGATCAACCAACGAAAAAC	1560
QY	1561	ttataatgatgcacatgattgtcacaagaatgacgggggcgcataccagtgtcttgtgcgaag	1620
Db	1561	TTATAAATGGAATCACTATGTTCAAAAGTGACGGGGGCGCATACCAAGTGTCTTGTGCCCAAGG	1620
QY	1621	acaaagctgcgcgtccaagaactatgtgcaggtgtgctctgaaataagaaactcccaaatata	1680
Db	1621	ACAAAGCTGTCGCTCAAGACTATGTGCAGGTGTGCTCTTGAATGAATGAACTCCCAAAATTA	1680
QY	1681	ttctgcctcttaagtgaaaaaagtgtgtagctccagacagacccggtttccctctatgtgcacg	1740
Db	1681	TTTCTGCTCTTATGTAAGAAAGTGCTGTAGTCCAGACAGCCGGTTTCCCTATGTGTCAMCG	1740
QY	1741	tgaaggaagaacactttgtcccaagatacattgtgaacccctgtgagatagaataccggaattctcaag	1800
Db	1741	TGAAGGGAACACCTTTGCCACATGTACGTGAGACCTGTGACATGATACCCGATTTCTCAAGG	1800
QY	1801	gttgacagctcacccgcatcaagcagatgatcaagtcggaaggaggaactgtgtcagctaccgga	1860
Db	1801	GTGGCAGTCAACCCGATATAGCCCAATATCATCGTGGAGGGGAACCTGGTGTACTACTCTGA	1860
QY	1861	acatctccagctcccaaggtccggagacggggaggtctaccgctgtcacctgtgccaaacaactcgg	1920
Db	1861	ACATCTCCAGCTCCACAGTCCGAGTCCGGAGCGGGAGGTCTACCGCTGTCACTGCCAAACACTCGG	1920
QY	1921	cgggaagtcgtctctgtacacagctcgaataaagcttaagagggcctcgaagcatctgaacca	1980
Db	1921	CGGAGTCTGTCCTGTACCAAGCTCTGAAATAACCTTAAGAGGGCTCTCAAGCATTCGACCAA	1980
QY	1981	tgaaaaacatcaacagcaatagcagaagccgggagaacatacattcaactgtcgtgtgaattgct	2040
Db	1981	TGAAAAACATCAACAGCAATAGCAGAGCGGAGACATACATTCACATGTGCTGTGATTTGCT	2040
QY	2041	atccgcttatctccatataaattgtataaagaactctaacctgtcttcccttcaaccaacggcc	2100
Db	2041	ATCCGCTATCTCCATTAATGTGTACAAACACTTAACCTGCTTCTTTCAACCAACCGCC	2100
QY	2101	aagtcgcatcttgagaacaacttgaaactcttaaaacttccaatltgcgaagaagatgtagcg	2160

QY	1021	atggtattggtataactaacctgcgtctgcatcagcggaatcgcataacacccggagaagacagagcgaga	1080
Db	1021	ATGGATTGTATTACTACACCGCTGCATACCGCGGCATCGATACACCGGAAGACGAGCGACGA	1080
QY	1081	gcaacagcgcgaagactctttgtaatacagaccagcgaaactcagccccatccatactcgtatg	1140
Db	1081	GCACACAGCCGACAGCTTTTGTATACGACCAGCCAGACTCACCCCCATTCATACGTGATG	1140
QY	1141	ggtttgcacatcgcaaaagccatagcctgaggcagcgtgtgtagcctgtcccttgcaaaagcgtcg	1200
Db	1141	GGTTTGACCAATCGCAAAAGCCATGGCTGGGACAGCGTGTGGAGCTGCTTGCAAAAGCGCTCG	1200
QY	1201	ggagcccttgagccagaattaccgcgtgtgccttaagaagaaacatgcccccttgtaactttcaggga	1260
Db	1201	GGCACCTTGAGCCAGATTACCGCTGCTGACTAGAGACAACTGCCCCCTTGAACTTTCAAGGGA	1260
QY	1261	ggttcacgaagaccgttgaacgggctgtcattatgtagaacaatlcgcgccctcggactcagaaga	1320
Db	1261	GGTTTCGAGAAGACCGGTGACGGGGCTGCTCATTTGAGAACATTGGCCCTCGGACTCAGGCA	1320
QY	1321	gctatgtcttcttgaaagtgtgtccaaacagatacggaaactgctaagggtgataaggccgctgtacg	1380
Db	1321	GCATGTGTTTGTGAAGTGTCCAAACAGATACGGAACGTCTAAGGTGATAGGCCGCTGTATCG	1380
QY	1381	tgaaacagccacattgaagaagccaccatcaatcccccagaagaagttaaaagcagctgtggtatcc	1440
Db	1381	TGAACACGCCACTGAAGAAGCCACCATCATGTCACAGAGAGTTAAAGCAGCGTGGGTACCC	1440
QY	1441	aagttccctgtgctcgtcgjgtagacagaacttgagaccagaactctcctgtgataccga	1500
Db	1441	AAGTTTCTTGTCTCTCTGTGAGGTGTCACGAGAACTGAGACAGGAACTCTCTGTATCCGCA	1500
QY	1501	atgtgtgaatccctccaaacccctcggaaaaaagtgtagatatacagaaggatacacaacagaaaac	1560
Db	1501	ATGTTGAATTCCTCAACCTCGGAATAAAATGTAGATATCACAGGATCAACCAACGAAAAAC	1560
QY	1561	ttataatgatcatcatatgttcaaaagtacacgggggcgcataccagtgtcttgtgcgcaag	1620
Db	1561	TTATATAGGATATCATATGTTCAAAAGTGACGGGGGGCCATACACAGTGTCTTGTGCCCAAG	1620
QY	1621	acaaagctgcgcgtccaagaactatgtgcaggtgtgctctgaagaatgaagactcccaaatata	1680
Db	1621	ACAAAGCTGTCGCTCAAGACTATGTACAGGTGTCTCTTGAATGTGAATCTCCCAAAATTA	1680
QY	1681	ttctgcgtcttaqtgaaaaaagtgtgtagtcacagacagacgggtttccctatgtgtcacag	1740
Db	1681	TTTCTGCTTTAGTGAATAAGTGCTGTAGTCCAGACAGCCGGTTTCCCTATGTGTCAAG	1740
QY	1741	tgaagggaacaaccttlycccagaatacatttgtaaccccttgagagataaacccgatttctcaag	1800
Db	1741	TGAAGGGAACAACCTTTTGCACAGATACGTGTGACCCCTGGACATGATACCCGATTCTCAAG	1800
QY	1801	gttgacagtcacccgatataagccagatgatcaagtcggaagggaagcttggtcagatcacctga	1860
Db	1801	GTGGCAGTCAACCCGATATAGCCCAATATCATCGTGGAGGGGAACCTGGTGTACTACTCTGA	1860
QY	1861	acatctccagctcccaaggtccggagacggggaggtctaacgcgtgcacatgcgccaaacactcgg	1920
Db	1861	ACATCTCCAGCTCCACAGTCCGAGTCCGGAGCGGGAGTCTACCGCTGCACCTGCCAAACACTCGG	1920
QY	1921	cgggaagtcgtctctgttaccagagctcgaataaagcttaagaaggagcctcgaagcatctgaacaa	1980
Db	1921	CGGAGTCTGTCCTGTACCAAGCTCTGAAATAACCTTAAGAGGGCTTCGAAGCATTCGACCAA	1980
QY	1981	tgaaaaacatcaacagcaatagcagaagccgggaacacataatactcgtcgtgtgatatggtct	2040
Db	1981	TGAAAAACATCAACAGCAATAGCAGAGCGGAGACATACATTCACATGTCGTGTGATTGGCT	2040
QY	2041	atccgcttatctccatataaagtgtataaagaactctaacctgtcttcccttcaacacccgccc	2100
Db	2041	ATCCGCTATCTCCATTAATGTGTACAAACACTTAACCTGCTTCTTTCAACCAACCGCCC	2100
QY	2101	aagtcgcatcttgagaacaacttgaaactccttaaaacttccaatltgcgaagaagatgtagcgt	2160

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 15:14:16 : Search time 216.06 Seconds
(without alignments)
3659.408 Million cell updates/sec

Title: US-08-956-991-1

Perfect score: 6604
Sequence: 1 tgactgagcgagcgacg.....gaaatgcgaataatatt 6604

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PC10S9.COMB.seq:*
7: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	541	8.2	1493	4	US-08-752-307B-6 Sequence 6, Appli
2	66.4	1.0	6000	1	US-08-348-006B-6 Sequence 6, Appli
3	66.4	1.0	6000	3	US-08-800-825A-6 Sequence 6, Appli
4	66.4	1.0	6000	6	PCT-US94-10166-6 Sequence 6, Appli
5	63	1.0	3507	4	US-08-775-009-36 Sequence 36, Appli
6	62.2	0.9	801	3	US-08-770-379-16 Sequence 16, Appli
7	62	0.9	4257	3	US-08-690-473-1 Sequence 1, Appli
8	62	0.9	12001	2	US-08-458-568A-11 Sequence 11, Appli
9	61.6	0.9	4403	3	US-08-284-941-1 Sequence 1, Appli
10	61.6	0.9	4403	6	US-08-447-642-1 Sequence 1, Appli
11	61.6	0.9	4403	6	PCT-US93-02147A-1 Sequence 1, Appli
12	60.4	0.9	8438	1	US-07-945-283-1 Sequence 1, Appli
13	60.2	0.9	1026	1	US-07-975-526-6 Sequence 6, Appli
14	59.4	0.9	4257	3	US-08-690-473-1 Sequence 1, Appli
15	59.4	0.9	12001	2	US-08-458-568A-11 Sequence 11, Appli
16	58.8	0.9	44377	4	US-08-804-198-1 Sequence 7, Appli
17	58.8	0.9	44377	4	US-08-804-198-1 Sequence 1, Appli
18	57.2	0.9	1548	4	US-08-763-106-5 Sequence 5, Appli
19	57.2	0.9	1581	4	US-08-763-106-5 Sequence 6, Appli
20	56.6	0.9	2823	1	US-08-398-008A-1 Sequence 1, Appli
21	56.6	0.9	2823	4	US-08-893-333-1 Sequence 1, Appli
22	56	0.8	1611	5	US-08-909-742-2 Sequence 2, Appli
23	55.6	0.8	43280	3	US-08-804-227C-1 Sequence 1, Appli
24	55.4	0.8	803	3	US-07-928-611-12 Sequence 12, Appli
25	55.4	0.8	803	3	US-08-487-811A-12 Sequence 12, Appli
26	55.4	0.8	803	6	PCT-US93-07370-12 Sequence 12, Appli
27	55.4	0.8	1610	1	US-08-056-051-5 Sequence 5, Appli

C 28	55.4	0.8	1610	1	US-07-928-611-21	Sequence 21, Appli
C 29	55.4	0.8	1610	3	US-08-487-811A-21	Sequence 21, Appli
C 30	55.4	0.8	1610	6	PCT-US93-07370-21	Sequence 21, Appli
C 31	55.2	0.8	2150	4	US-08-318-837-1	Sequence 1, Appli
C 32	55.2	0.8	5117	6	PCT-US95-05512-1	Sequence 1, Appli
C 33	55	0.8	936	1	US-08-018-97C-4	Sequence 4, Appli
C 34	54.6	0.8	1053	7	5352575-6	Patent No. 5352575
C 35	54.2	0.8	1550	3	US-08-609-443B-17	Sequence 12, Appli
C 36	54.2	0.8	5467	1	US-07-745-206A-12	Sequence 12, Appli
C 37	54.2	0.8	5467	1	US-08-311-363-12	Sequence 12, Appli
C 38	54.2	0.8	7175	2	US-08-455-543A-8	Sequence 8, Appli
C 39	54.2	0.8	7175	3	US-08-193-078B-8	Sequence 8, Appli
C 40	54.2	0.8	7175	3	US-08-223-305C-8	Sequence 8, Appli
C 41	54.2	0.8	7175	3	US-08-149-097D-8	Sequence 8, Appli
C 42	54.2	0.8	7362	2	US-08-455-543A-7	Sequence 7, Appli
C 43	54.2	0.8	7362	3	US-08-193-078B-7	Sequence 7, Appli
C 44	54.2	0.8	7362	3	US-08-223-305C-7	Sequence 7, Appli
C 45	54.2	0.8	7362	3	US-08-149-097D-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-752-307B-6

: Sequence 6, Application US/08752307B
: Patent No. 5952171

: GENERAL INFORMATION:

: APPLICANT: McCarthy, Sean A.

: APPLICANT: McCarthy, David P.

: APPLICANT: Levinson, Douglas A.

: TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES

: NUMBER OF SEQUENCES: 14

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Fish & Richardson, P.C.

: STREET: 225 Franklin Street

: CITY: Boston

: STATE: MA

: COUNTRY: US

: ZIP: 02110-2804

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: Windows95

: SOFTWARE: FastSeq for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/752,307B

: FILING DATE: 19-NOV-1996

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER:

: FILING DATE:

: ATTORNEY/AGENT INFORMATION:

: NAME: Melkijohn, Ph.D., Anita L.

: REGISTRATION NUMBER: 35,283

: REFERENCE/DOCKET NUMBER: 09404/020001

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 617-542-5070

: TELEFAX: 617-542-8906

: TELEX: 200154

: INFORMATION FOR SEQ ID NO: 6:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1493 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: cDNA

: FEATURE:

: NAME/KEY: Coding Sequence

: LOCATION: 99...1493

: US-08-752-307B-6

C	3	397.4	6.0	419	42	A1094516	A1094516	oy63f11.s
C	4	374.2	5.7	510	25	N80326	N80326	y291f03.r1
C	5	358.8	5.4	430	35	N64532	N64532	y291f03.s1
C	6	343.8	5.2	355	37	AJ003472	AJ003472	AJ003472
C	7	341.2	5.2	458	46	A154704	A154704	UI-R-BRO-
C	8	318	4.8	321	20	Z41519	Z41519	HSC2XW042 n
C	9	303.4	4.6	307	41	A1051251	A1051251	oy49e04.x
C	10	288	4.4	310	21	F13426	F13426	HSC2XW041 n
C	11	278	4.2	294	20	Z45894	Z45894	HSC2XW041 n
C	12	271.2	4.1	427	71	AM159035	AM159035	z45e05.x
C	13	262.4	4.0	365	64	AM048531	AM048531	UI-M-BH1-
C	14	254.8	3.9	337	29	AM170935	AM170935	ms45h05.r
C	15	242.4	3.7	252	21	F09100	F09100	HSC2XW022 n
C	16	213.4	3.2	254	46	A1448860	A1448860	ms45h03.x
C	17	187	2.8	187	39	AA878733	AA878733	oe80d01.s
C	18	175.6	2.7	619	79	FR0002941	FR0002941	F.rubripes
C	19	163.2	2.5	402	64	AM048129	AM048129	UI-M-BH1-
C	20	156	2.4	156	62	A1880033	A1880033	ap24b07.x
C	21	147.4	2.2	409	63	AM014717	AM014717	UI-H-B10-
C	22	143	2.2	147	79	HSWC18B12	HSWC18B12	X88325 H.sapiens D
C	23	142.4	2.2	465	81	B07180	B07180	GSRU72-h3-u
C	24	138.4	2.1	450	47	A1498283	A1498283	tm85h11.x
C	25	123	1.9	646	42	A1106884	A1106884	GH06134.5
C	26	122.8	1.9	602	79	FR0022038	FR0022038	F.rubripes
C	27	120.6	1.8	442	79	FR0030733	FR0030733	AL027102 Fugu rubr
C	28	118.4	1.8	437	79	FR0002934	FR0002934	Z86717 F.rubripes
C	29	113.4	1.7	298	31	AA325603	AA325603	EST28915
C	30	111.4	1.7	315	104	AO552303	AO552303	RPC1-11-4
C	31	105.4	1.6	935	79	CNS006XK	CNS006XK	AL056051 Drosophila
C	32	104.6	1.6	282	71	AV264594	AV264594	AL066742 Drosophila
C	33	104	1.6	932	79	CNS010RX	CNS010RX	AL099332 Drosophila
C	34	98.6	1.5	776	79	CNS006XK	CNS006XK	AL066051 Drosophila
C	35	97.4	1.5	735	79	CNS01523	CNS01523	AL019091 Drosophila
C	36	95	1.4	1101	80	AO893056	AO893056	HS_4832-A
C	37	94.8	1.4	1080	88	AO893056	AO893056	AL010949 Drosophila
C	38	94	1.4	1101	80	CNS0153F	CNS0153F	AL053013 Drosophila
C	39	93.6	1.4	925	79	CNS0091P	CNS0091P	AL018162 Drosophila
C	40	93.6	1.4	1035	80	CNS017K0	CNS017K0	X88135 H.sapiens D
C	41	92.6	1.4	925	79	HSWC34D09	HSWC34D09	AL030331 Drosophila
C	42	92.6	1.4	925	79	CNS0091P	CNS0091P	AL014951 F.rubripes
C	43	92.2	1.4	544	79	FR0022080	FR0022080	AL066742 Drosophila
C	44	91.8	1.4	932	79	CNS00720	CNS00720	AL066742 Drosophila
C	45	91.4	1.4	894	88	AO893147	AO893147	HS_3131-B

ALIGNMENTS

RESULT 1
A1243628/c 439 bp mRNA EST 01-DEC-1998
LOCUS qh88a03.x1 Soares_NFL_T_GBC-S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1854028 3', mRNA sequence.
ACCESSION A1243628
VERSION A1243628.1 GI:3839025
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 439)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043430.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 476 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source

Location/O.alifiers

1. 439

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1854028"

/clone_11b="Soares_NFL_T_GBC-S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pRT3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI-CCAP-GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the s we 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT

86 a

104 c

115 g

133 t

1 others

ORIGIN

Query Match

Best Local Similarity 98.1%; Pred. No. 4,1e-70;

Matches 420; Conservative

Mismatches 8; Indels 0; Gaps 0;

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct

6177 ttttgaacagagagtggtccagcaccagcagagctagctagctagctagct


```

195 YRCLTIRHRTGETRQSAARLFWSDPANSAFSLDDGDKKAAAGORVLEPCALALHPEP 254
241 DYRLKDNMBLELSGRFOKTVTGTLLENIRPDSGSGYCEVSNRYGTAKVIGRLYKQPL 300
255 DYRLKDNMBLELSGRFOKTVTGTLLENIRPDSGSGYCEVSNRYGTAKVIGRLYKQPL 314
301 KATISPRKVVSSVSGOVSLSCSVTGTEDOLSWYRNGEILNPKNRTIGCINHEMLIMH 360
315 KATISPRKVVSSVSGOVSLSCSVTGTEDOLSWYRNGEILNPKNRTIGCINHEMLIMH 374
361 MVSMDGAYOCFARKDKLSADYVQVUVEEDGTPKIIISAFSEKVVSPAEPYSLMCNKGTP 420
375 MVSMDGAYOCFARKDKLSADYVQVUVEEDGTPKIIISAFSEKVVSPAEPYSLMCNKGTP 434
421 LPTITWTLDDPILKGGSHRISOMITSEGNVSYLNISSQVBDGCVYRCTANNSAGVVL 480
435 LPTITWTLDDPILKGGSHRISOMITSEGNVSYLNISSQVBDGCVYRCTANNSAGVVL 494
481 YQARIINRGPAISIRPMKNITAIAGROTYIHCRVYGYPSYIKMKNSNLLPENHROVAFE 540
495 YQARIINRGPAISIRPMKNITAIAGROTYIHCRVYGYPSYIKMKNSNLLPENHROVAFE 554
541 NNGTLKLSDVQKEVEDEGEYCNVLYQPOLSTSQSVHVTVPPEFIOFEPFRPSIGORVF 600
555 NNGTLKLSDVQKEVEDEGEYCNVLYQPOLSTSQSVHVTVPPEFIOFEPFRPSIGORVF 614
601 IPCVYVSGDLPIITTTOKODRAPIPGSLGVTIIDNIDFTSSLRISNLSLMHNGNTTCIARNE 660
615 IPCVYVSGDLPIITTTOKODRAPIPGSLGVTIIDNIDFTSSLRISNLSLMHNGNTTCIARNE 674
661 AAAREHOSOLIVRPKPFVVOOPRODGIYKAVILNCSAGIEPPTIVMKFSGAGVPOF 720
675 AAAREHOSOLIVRPKPFVVOOPRODGIYKAVILNCSAGIEPPTIVMKFSGAGVPOF 734
721 OPALNRIQVLSNGSLIKHVEEDSGYLLCKVSNVDGADVSKMYLTVKIPAMITSTYP 780
735 OPALNRIQVLSNGSLIKHVEEDSGYLLCKVSNVDGADVSKMYLTVKIPAMITSTYP 794
781 NTTLATOGQKKEMJCTAHGEKPIIVRWKEDRIINPEMARIVLSTKVEGEVISTLOILP 840
795 NTTLATOGQKKEMJCTAHGEKPIIVRWKEDRIINPEMARIVLSTKVEGEVISTLOILP 854
841 TVRDSGFFSCHALNSYGEDRGITQLTVQEPDPPEIEIKDVAKRTITLWMTMGFDGNSP 900
855 TVRDSGFFSCHALNSYGEDRGITQLTVQEPDPPEIEIKDVAKRTITLWMTMGFDGNSP 914
901 ITGYDIECKNKSQSDMSAQRTKDVSPOLNSATIIDIHPSSTYSIRMYAKNRIGKSEPSNE 960
915 ITGYDIECKNKSQSDMSAQRTKDVSPOLNSATIIDIHPSSTYSIRMYAKNRIGKSEPSNE 974
961 LTIADAEAPRGPOEYHLEPPISSOSIRVTKAPKAKHLONGIIRNGYIGREVSTGNGFO 1020
975 LTIADAEAPRGPOEYHLEPPISSOSIRVTKAPKAKHLONGIIRNGYIGREVSTGNGFO 1034
1021 FNISVDTSGSEYVITDNLNKKFTQYGLVQACNRAGTGSSEIITTTLEDVPSYPEN 1080
1035 FNISVDTSGSEYVITDNLNKKFTQYGLVQACNRAGTGSSEIITTTLEDVPSYPEN 1094
1081 VQAIATSPESIS:SMSTLSKALNGILOGFRVIYMANLMGELGELKINITTQPSLELDG 1140
1095 VQAIATSPESIS:SMSTLSKALNGILOGFRVIYMANLM GELGELIN:ITQPSLELDG 1154
1141 LEKTNVSIQVLAFTRAGDGVRSQOIFRTKEDVPGPPAGVKAASASAMFVSWLPLK 1200
1155 LEKTNVSIQVLAFTRAGDGVRSQOIFRTKEDVPGPPAGVKAASASAMFVSWLPLK 1214
1201 LNGIIRKVTVCSPHPYVISEFASPDPSFYRIPLNLSRNQOYSVWVAVYTSAGRGSSE 1260
1215 LNGIIRKVTVCSPHPYVISEFASPDPSFYRIPLNLSRNQOYSVWVAVYTSAGRGSSE 1274
1261 ITTVEPLAKAARILITTSYGYTTTMMKDIYVPCAVAGDPSP/ VMMKDSNGTSPSLVTIDG 1320

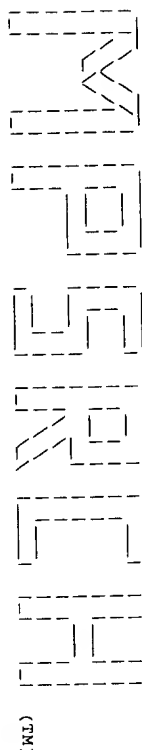
```

```

QY 1275 ITTVEPLAKAARILITTSYGYTTTMMKDIYVPCAVAGDPSPAVVMMKDSNGTSPSLVTIDG 1334
DB 1321 RRSITFNSGFIITRTVKAEDSGYIYCIANNMGSDEITLNIQVYPPPOPLUYSKTSSS 1380
QY 1335 RRSITFNSGFIITRTVKAEDSGYIYCIANNMGSDEITLNIQVYPPPOPLUYSKTSSS 1394
DB 1381 ITLSWLPDNGSGSIRGIILOYSEDNSBOWGSPISSESYRLLENKCGTWKFKTLTAQ 1440
QY 1395 ITLSWLPDNGSGSIRGIILOYSEDNSBOWGSPISSESYRLLENKCGTWKFKTLTAQ 1454
DB 1441 NGVGRISSEIIEAKTLCKEPOFSKEQELFASINTVRNLNIGMNDGGCPITSFTLEYR 1500
QY 1455 NGVGRISSEIIEAKTLCKEPOFSKEQELFASINTVRNLNIGMNDGGCPITSFTLEYR 1514
DB 1501 PEGTIVWTTAQTSTLSKSYILYDLOEATWYELQVRVNSAGACAKQANFATLNYDGTIP 1560
QY 1515 PEGTIVWTTAQTSTLSKSYILYDLOEATWYELQVRVNSAGACAKQANFATLNYDGTIP 1574
DB 1561 PLIKSVQNEGLTTEGKMLVTISCLVGVLLFVLLVYRRRRRQRLKRLDQAKSL 1620
QY 1575 PLIKSVQNEGLTTEGKMLVTISCLVGVLLFVLLVYRRRRRQRLKRLDQAKSL 1634
DB 1621 AEMLSKNTRTSDTSLSKOQOTLRMHIDI PRAQLLIEERDTEMIDDSYSTVLLDADGFA 1680
QY 1635 AEMLSKNTRTSDTSLSKOQOTLRMHIDI PRAQLLIEERDTEMIDDSYSTVLLDADGFA 1694
DB 1681 AKOKSLTHTVHOSVSOATGPLVDSDARPPTTRAKAGPARRNRYASQWTLNR 1740
QY 1695 AKOKSLTHTVHOSVSOATGPLVDSDARPPTTRAKAGPARRNRYASQWTLNR 1754
DB 1741 PAPTISATHTLTDMRLPTPRAAGSYDKESDYSVSPQDTRARRSNVSTESASTYEEL 1800
QY 1755 PAPTISATHTLTDMRLPTPRAAGSYDKESDYSVSPQDTRARRSNVSTESASTYEEL 1814
DB 1801 ARAYEHA :MEQOLRAKATTIECFISDTSSQULAGTNEYIDLSITSPSSGICREFTAS 1860
QY 1815 ARAYEHA :MEQOLRAKATTIECFISDTSSQULAGTNEYIDLSITSPSSGICREFTAS 1874
DB 1861 PPKPODGGRYVMNAVPAKAIQVTSYICLHTEWTF 1896
QY 1875 PPKPODGGRYVMNAVPAKAIQVTSYICLHTEWTF 1910

RESULT 2
ENTRY 158164 #type complete
TITLE BIG-1 protein - rat
ORGANISM #formal_name Kallos norvegicus #common_name Norway rat
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1996
ACCESSIONS 158164
REFERENCE 158164
AUTHORS Yoshihara, Y.; Kawasaki, M.; Tani, A.; Tamada, A.; Nagata, S.; Kagamiyama, H.; Mori, K.
#journal Neuron (1994) 13:415-426
#title BIG-1: a new TAG-1/f3-related member of the immunoglobulin superfamily with neurite outgrowth-promoting activity.
#cross-references MUID:94338697
#accession 158164
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1028 #label RES
#cross-references EMBL:U1031; NID:g563132; PIDN:AAA65607.1; PID:g563133
GENETICS
#gene BIG-1
CLASSIFICATION #superfamily contactin; fibronectin type III repeat homology; immunoglobulin homology
SUMMARY #length 1028 #molecular_weight 112788 #checksum 5866
Query Match 6.0%; Score 809; DB 2; Length 1028;
Best Local Similarity 24.7%; Pred. No. 4,316-127;
Matches 250; Conservative 249; Mismatches 447; Indels 66; Gaps 54;

```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Apr 25 12:34:34 2000; Maspar time 57.44 Seconds
993.145 Million cell updates/sec

Tabular output not generated.

Title: >US-08-956-991-2
Description: (1.1910) from US08956991A.pep
Perfect Score: 13516
Sequence: 1 MWIALSLFQSFANVSEDL.....AIGQVSTYICHTLEMTFC 1910

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 58.721; Variance 101.190; scale 0.580
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	744	5.5	1447	1	DCC_MOUSE TUMOR SUPPRESSOR PROTE	3.56e-133
2	740	5.5	1447	1	DCC_HUMAN TUMOR SUPPRESSOR PROTE	3.04e-132
3	720	5.3	1040	1	AXOI_HUMAN AXONIN-1 PRECURSOR (AX	1.35e-127
4	703	5.2	1036	1	AXOI_CHICK AXONIN-1 PRECURSOR	1.18e-123
5	701	5.2	1040	1	AXOI_RAT AXONIN-1 PRECURSOR (AX	3.43e-123
6	677	5.0	1284	1	NRCA_CHICK NG-CAM RELATED CELL AD	1.22e-117
7	663	4.9	1010	1	CONTRACTIN PRECURSOR (N	2.06e-114
8	631	4.7	1018	1	CONTRACTIN PRECURSOR (C	4.68e-107
9	631	4.7	1020	1	CONTRACTIN PRECURSOR (N	4.97e-108
10	637	4.7	1259	1	CAML_RAT NEURAL CELL ADHESION M	1.97e-108
11	637	4.7	1260	1	CAML_MOUSE NEURAL CELL ADHESION M	1.55e-104
12	620	4.5	1257	1	CAML_HUMAN NEURAL CELL ADHESION M	5.09e-102
13	609	4.5	2029	1	LAR_DROME LAR PROTEIN PRECURSOR	1.65e-99
14	598	4.4	1897	1	PIPF_HUMAN NEURAL CELL ADHESION M	3.87e-93
15	570	4.2	837	1	NCM2_MOUSE NEURAL CELL ADHESION M	2.52e-91
16	562	4.2	837	1	NCM2_HUMAN NEURAL CELL ADHESION M	2.52e-91
17	562	4.2	1239	1	NRG_DROME NEUROGLIAN PRECURSOR	4.79e-94
18	574	4.2	1912	1	PTPD_HUMAN NEURAL CELL ADHESION M	2.03e-90
19	558	4.1	1091	1	NCAL_CHICK NEURAL CELL ADHESION M	4.99e-87
20	543	3.9	1088	1	NCAL_HUMAN NEURAL CELL ADHESION M	9.48e-83
21	524	3.9	1088	1	NCAL_XENLA NEURAL CELL ADHESION M	1.59e-82
22	533	3.9	1092	1	NCAL_XENLA NEURAL CELL ADHESION M	1.59e-82
23	531	3.9	1266	1	NGCA_CHICK NEUROVAL-GLIAL CELL AD	2.53e-84

24	509	3.8	761	1	NCAL_HUMAN NEURAL CELL ADHESION M	2.20e-79
25	498	3.7	853	1	NCAL_BOVIN NEURAL CELL ADHESION M	6.33e-77
26	496	3.7	1115	1	NCAL_MOUSE NEURAL CELL ADHESION M	1.77e-76
27	492	3.6	858	1	NCAL_RAT NEURAL CELL ADHESION M	1.38e-75
28	478	3.5	725	1	NCAL_MOUSE NEURAL CELL ADHESION M	1.80e-72
29	401	3.0	898	1	FAS2_SCHAM FASCICLIN II PRECURSOR	1.41e-55
30	402	3.0	3707	1	EGBM_MOUSE BASEMENT MEMBRANE-SPEC	8.57e-56
31	378	2.8	811	1	FS21_DROME FASCICLIN II, MEMBRANE	1.26e-50
32	378	2.8	873	1	FS21_DROME FASCICLIN II, MEMBRANE	1.26e-50
33	364	2.7	4393	1	PGBM_HUMAN BASEMENT MEMBRANE-SPEC	6.10e-37
34	313	2.3	2477	1	FINC_RAT FIBRONECTIN PRECURSOR	6.10e-36
35	308	2.3	2477	1	FINC_MOUSE FIBRONECTIN PRECURSOR	6.10e-36
36	313	2.3	2481	1	UN52_CAEEL BASEMENT MEMBRANE PROT	6.10e-37
37	294	2.2	2265	1	FINC_BOVIN FIBRONECTIN (FN)	4.83e-33
38	287	2.1	345	1	OPCM_BOVIN OPTIC BINDING PROTEIN	1.27e-31
39	284	2.1	345	1	OPCM_HUMAN OPTIC BINDING PROTEIN	5.14e-31
40	279	2.1	345	1	OPCM_RAT OPTIC BINDING PROTEIN	5.22e-30
41	286	2.1	1131	1	MYPE_CHICK MYOSIN-BINDING PROTEIN	2.03e-31
42	264	2.0	337	1	G55A_CHICK NEURITE INHIBITOR GP55	5.13e-27
43	267	2.0	1142	1	MYPE_HUMAN MYOSIN-BINDING PROTEIN	1.30e-27
44	262	1.9	333	1	AMAL_DROME AMALGAM PROTEIN PRECUR	1.28e-26
45	262	1.9	2386	1	FINC_HUMAN FIBRONECTIN PRECURSOR	1.28e-26

ALIGNMENTS

RESULT 1
ID DCC_MOUSE STANDARD: PRT: 1447 AA.
AC P70211:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.
DCC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE: 96112625.
RA COOPER H.M., ARMES P., BRITTO J., GAD J., WILKS A.F.;
RT "Cloning of the mouse homologue of the developing cancer
gene (mDCC) and its expression in the developing mouse embryo."
RN Oncogene 11:2243-2254(1995).
RN [2]
RP REVISIONS.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RA COOPER H.M.;
RL Submitted (JUN-1996) to the IMHJ/Genbank/DBJ databases.
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD
FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION, HIGHEST LEVELS
EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
AND REMAIN AT THIS LEVEL IN THE ADULT.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).


```

|||||
QY 255 DYRWMLKDNMPLELSGRFQKTVTGLLENIIRPSDGSYCEVSNRGTAKVIGRLTVKQPL 314
Db 301 KATISPRKXSSVGSQVSLSCSVTGTEDDELSTWYRNGELNPKGNVTRITGTHHEMLMDH 360
QY 315 KATISPRKXSSVGSQVSLSCSVTGTEDDELSTWYRNGELNPKGNVTRITGTHHEMLMDH 374
Db 361 MKSDGAYQCFVRKDKLSAQDYVQVVLDEDTPKIISAFSEKVPSPAEVSLMCVWKGTP 420
QY 375 MKSDGAYQCFVRKDKLSAQDYVQVVLDEDTPKIISAFSEKVPSPAEVSLMCVWKGTP 434
Db 421 LPTITWTLDDDDPILKGGSHRISQMITSEGNVSYLNISSSQV?CGVYRGTANNSAGVYL 480
QY 435 LPTITWTLDDDDPILKGGSHRISQMITSEGNVSYLNISSSQV?CGVYRGTANNSAGVYL 494
Db 481 YQARINVRGPASIRPKMTALAGRDYTHCRVIGPYYSIKWYNSNLLPBNHNOVAFE 540
QY 495 YQARINVRGPASIRPKMTALAGRDYTHCRVIGPYYSIKWYNSNLLPBNHNOVAFE 554
Db 541 NNGTLKLSQKVEDEGEYTCNVLPOLSTSQSVHVTYKVPFIQPEEPFESIGQVRF 600
QY 555 NNGTLKLSQKVEDEGEYTCNVLPOLSTSQSVHVTYKVPFIQPEEPFESIGQVRF 614
Db 601 PCVWVSGDLPTITWQKGRPIPGSLGYTIDNIDFTSSLRSNLSLHNGNYTCIARNE 660
QY 615 PCVWVSGDLPTITWQKGRPIPGSLGYTIDNIDFTSSLRSNLSLHNGNYTCIARNE 674
Db 661 AAAYEHOSOLIVRPKPEYVQPRDQDGIKAVILNCSAEGYPVTIYMKFSKAGVQF 720
QY 675 AAAYEHOSOLIVRPKPEYVQPRDQDGIKAVILNCSAEGYPVTIYMKFSKAGVQF 734
Db 721 OPALNGRIQVLSNGLIKHVEEDSGYLLCKVSNVGYADVSKSMYLTVKIPAMITSYP 780
QY 735 OPALNGRIQVLSNGLIKHVEEDSGYLLCKVSNVGYADVSKSMYLTVKIPAMITSYP 794
Db 781 NTTLATOGKCKKMSCTARHEKPIYVRMEKEDIIMPEAR:IVTSKEVEEYISTLOLTP 840
QY 795 NTTLATOGKCKKMSCTARHEKPIYVRMEKEDIIMPEAR:IVTSKEVEEYISTLOLTP 854
Db 841 TVREDSGFFSCHAINS YGEGDRCIIQIIVQEPDPPEIEIKDKVARTITLMTMGFDGNSP 900
QY 855 TVREDSGFFSCHAINS YGEGDRCIIQIIVQEPDPPEIEIKDKVARTITLMTMGFDGNSP 914
Db 901 ITGYDIECKNKSQSDWSAORTKDVSPOLNSATIIDIHPSSTYSIMVAKNRIGSEPSNE 960
QY 915 ITGYDIECKNKSQSDWSAORTKDVSPOLNSATIIDIHPSSTYSIMVAKNRIGSEPSNE 974
Db 961 LTTTADDEAPDPGPQEVHLEPISSOSIRVTWKAPKKHILQNGIIRGYOIGREYSTGGMFQ 1020
QY 975 LTTTADDEAPDPGPQEVHLEPISSOSIRVTWKAPKKHILQNGIIRGYOIGREYSTGGMFQ 1034
Db 1021 FNII SVDRSGDSEVYTLDLNKFQYGLVQVACNAGTGPSSQELITTTLEDVPSYPEN 1080
QY 1035 FNII SVDRSGDSEVYTLDLNKFQYGLVQVACNAGTGPSSQELITTTLEDVPSYPEN 1094
Db 1081 VOALATSPESISISWSTLSKEALNGLQCFRVIYVWNLMDGELGKINITTQPSLEIDG 1140
QY 1095 VOALATSPESISISWSTLSKEALNGLQCFRVIYVWNLMDGELGKINITTQPSLEIDG 1154
Db 1141 LEKYNYSIOVLAFTRADGVSSEIOIFTRTKEDVGPAGVAKAAASASMFVSLPPLK 1200
QY 1155 LEKYNYSIOVLAFTRADGVSSEIOIFTRTKEDVGPAGVAKAAASASMFVSLPPLK 1214
Db 1201 LKNGIIRKTYVFCSHPYPTVISEFEASPDSEFYSIRIPNLSRN:QYSVWVAVVSAGKGNSE 1260
QY 1215 LKNGIIRKTYVFCSHPYPTVISEFEASPDSEFYSIRIPNLSRNQYSVWVAVVSAGKGNSE 1274
Db 1261 IITVEPLAKAPARILITFSGTITVT?PMKQDILVPCKAVGDPSPAVKMKKSNCTP?SLVITDG 1320
QY 1275 IITVEPLAKAPARILITFSGTITVT?PMKQDILVPCKAVGDPSPAVKMKKSNCTP?SLVITDG 1334
Db 1321 RRSIFSNGSFIIRTVKAEDSGYSCIANNNMGSDIELNLQVQVPPDPRLTVSKTSSS 1380

```

```

QY 1335 RRSIFSNGSFIIRTVKAEDSGYSCIANNNMGSDIELNLQVQVPPDPRLTVSKTSSS 1394
Db 1381 IITSLMLPGDNGSSIRNGYIILQYSEDNSEQMGSPFIPSPERSRLEMLKCGTYWKTTLTAQ 1440
QY 1395 IITSLMLPGDNGSSIRNGYIILQYSEDNSEQMGSPFIPSPERSRLEMLKCGTYWKTTLTAQ 1454
Db 1441 NGVPGRISEIIEAKTLGKEPOFSKEQELFASINTRVRLNLIGWDDGCPITSFLEVR 1500
QY 1455 NGVPGRISEIIEAKTLGKEPOFSKEQELFASINTRVRLNLIGWDDGCPITSFLEVR 1514
Db 1501 PFGTTWTTAORTSLSKSYLLYDLOEATWYELQMRVCSAGCAEQANFATLNDGSTIP 1560
QY 1515 PFGTTWTTAORTSLSKSYLLYDLOEATWYELQMRVCSAGCAEQANFATLNDGSTIP 1574
Db 1561 PLIKSVONEEGTLTNEGKMLVTISCLLVGVLLFVLLLVVRRRRERLRLDASTL 1620
QY 1575 PLIKSVONEEGTLTNEGKMLVTISCLLVGVLLFVLLLVVRRRRERLRLDASTL 1634
Db 1621 AEMLSKNTRTSDTLKQOQTLRMHIDIPRAQLLIEERDTMETIDRSTVLLTDADFGFA 1680
QY 1635 AEMLSKNTRTSDTLKQOQTLRMHIDIPRAQLLIEERDTMETIDRSTVLLTDADFGFA 1694
Db 1681 AKOKSLTVHTVHYQSVQATGCLVDVSDARPGTNETTRNAKAGPTANRYASQWTLNR 1740
QY 1695 AKOKSLTVHTVHYQSVQATGCLVDVSDARPGTNETTRNAKAGPTANRYASQWTLNR 1754
Db 1741 PHPTISAHITLTTWRLPTPRAGSVNDESDSYSVSPQDTPDARSMTESASSTYEEL 1800
QY 1755 PHPTISAHITLTTWRLPTPRAGSVNDESDSYSVSPQDTPDARSMTESASSTYEEL 1814
Db 1801 ARAYEHAKMEQOLRHAKEFTTTECFISDTSSEQLTAGTNEYTSLTSTI SESEICRFTAS 1860
QY 1815 ARAYEHAKMEQOLRHAKEFTTTECFISDTSSEQLTAGTNEYTSLTSTI SESEICRFTAS 1874
Db 1861 PRKPODGGRYMNAVPAKAIQVTSYICLTLTLEWTCF 1896
QY 1875 PRKPODGGRYMNAVPAKAIQVTSYICLTLTLEWTCF 1910

RESULT 2
ID 060469 PRELIMINARY: PRT: 1571 AA.
AC 060469:
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE.
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA YAMAKAWA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
RA LYONS G.E., KORENBERG J.R.;
RL Submitted (sep-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF023450; AAC17967.1;
DR PIR: P00041; f03: 6.
DR PIR: P00047; i07: 8.
DR PRINTS: P00014; FNTYPEIII.
DR SEQUENCE 1571 AA: 5F8C77D1 CRC32:

Query Match 82.3%; Score 11119; DA 4; Length 1571;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1561; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:31:09 ; Search time 383.18 Seconds
(without alignments)
13.059 Million cell updates/sec

Title: US-08-956-991-5

Perfect score: 20

Sequence: 1 ccaggtctcaagaagcagcagc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 23170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance - you have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	1 V31983	Human Down syndrom
2	20	100.0	6413	1 V31988	Human Down syndrom
3	20	100.0	6604	1 V31981	Human Down syndrom
4	18.4	92.0	2173	1 V31987	Mouse Down syndrom
5	16.8	84.0	187	1 T22739	Human gene signatu
6	16.4	82.0	12923	1 N90338	Sequence of human
7	15.8	79.0	5406	1 V52299	Streptococcus pneu
8	15.4	77.0	4646	1 X20577	Poly nucleotide seq
9	15.2	76.0	3417	1 T39622	Human DNA ligase I
10	15.2	76.0	3417	1 V82491	Human DNA ligase I
11	15.2	76.0	8298	1 O50168	Human DNA ligase I
12	15.2	76.0	12752	1 O47380	p190 protein gene.
13	15.2	76.0	110000	1 T58840_1	Sequence which cor
14	14.8	74.0	2071	1 V46494	Continuation (2 of
15	14.8	74.0	2975	1 X03792	Interleukin-1 beta
16	14.8	74.0	8450	1 T74994	Human semaphorin E
17	14.8	74.0	8535	1 T74994	Plasmodium falciparum
18	14.8	74.0	9326	1 O25051	Galv SEMO genome.
19	14.4	72.0	709	1 X07563	DNA complementary
20	14.4	72.0	1185	1 V43608	Homo sapiens eosin
21	14.4	72.0	1537	1 V59804	Human secreted pro
22	14.4	72.0	1605	1 V59687	Human secreted pro
23	14.4	72.0	1871	1 O26959	Human secreted pro
24	14.4	72.0	2132	1 N60202	Guinea pig PAR rec
25	14.4	72.0	2132	1 T67151	Interferon-pseudo
26	14.4	72.0	2350	1 T61290	Arabidopsis CER1 g
27	14.4	72.0	3517	1 T61291	Mouse beta-3 integ
28	14.4	72.0	3534	1 T67410	Mouse truncated be
29	14.4	72.0	3549	1 T68137	H. pylori cytoplas
30	14.4	72.0	3648	1 O64580	120-128 kilodalton
31	14.4	72.0	3648	1 T14052	Helicobacter pylori
32	14.4	72.0	4149	1 V62461	Helicobacter pylori
33	14.4	72.0	4821	1 O64581	Tag A antigen of H
34	14.4	72.0	4821	1 O66728	H. pylori tagA ant

35	14.4	72.0	4821	1 T14051	Helicobacter pylori
36	14.4	72.0	5925	1 O48733	CAI gene. Recombin
37	14.4	72.0	19932	1 T46159	CagI locus. Helico
38	14.4	71.0	288	1 T60756	HIV-1 strain YBF30
39	14.2	71.0	560	1 T59027	PDD16b left juncti
40	14.2	71.0	580	1 T16308	Pea H74 cDNA parti
41	14.2	71.0	600	1 V21212	Human type II GNRH
42	14.2	71.0	600	1 V26007	Bacteriophage DNA
43	14.2	71.0	874	1 V59097	Clathrin-associat
44	14.2	71.0	1101	1 V59654	Human secreted pro
45	14.2	71.0	1422	1 O63870	Feline T cell prot

ALIGNMENTS

RESULT 1	
ID V31983	V31983 standard; cDNA; 20 BP.
AC V31983;	
DT 28-SEP-1998 (first entry)	
DE Human Down syndrome-cell adhesion molecule DS-CAM primer B9-131F.	
KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; PCR; primer: ss.	
OS Synthetic.	
OS Homo sapiens.	
PN WO9817795-A1.	
PD 30-APR-1998.	
PF 23-OCT-1997; U19547.	
PR 25-OCT-1996; US-029322.	
PA (CEDA-) CEDARS SINAI MEDICAL CENT.	
PI Korenberg JR;	
DR MPI: 98-271791/24.	
PT New isolated Down's Syndrome-cell adhesion molecule - used to	
PT develop products for detection, diagnosis and therapy of	
PT developmental and neurological abnormalities	
PS Example 5; Page 79; 109pp; English.	
CC Forward primer B9-131F and reverse primer B9-131R (see V31984)	
CC were used in RT-PCR assays of cDNA libraries of various human	
CC tissues to determine human Down syndrome-cell adhesion molecule	
CC DS-CAM (see V31981 and V31988) expression. The results	
CC demonstrated expression of human DS-CAM mRNA in foetal and adult	
CC brain, and foetal kidney. In addition, a breast carcinoma cell	
CC line showed expression of human DS-CAM mRNA. DS-CAM polypeptides	
CC (see W42086-87) are associated with developmental and neurological	
CC processes. The polypeptides and nucleic acids are used to develop	
CC and neurological abnormalities.	
SQ Sequence 20 BP: 5 A; 5 C; 6 G; 3 T;	
Query Match	100.0%; Score 20; DB 1; Length 20;
Best Local Similarity	100.0%; Pred. No. 0.41; Mismatches 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0;	
QY 1 ccaggtctcaagaagcagcagc 20	
DB 1 CCAGTCTCAAGAGCAGCAGC 20	
RESULT 2	
ID V31988	V31988 standard; cDNA; 6413 BP.
AC V31988;	
DT 28-SEP-1998 (first entry)	
DE Human Down syndrome-cell adhesion molecule DS-CAM2 cDNA.	
KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; ds; ss.	
OS Homo sapiens.	

FH Key Location/Qualifiers
 FT CDS 453..5168
 FT /*tag= a
 PN MO9817795-A1.
 PD 30-APR-1998.
 PF 23-OCT-1997: U19547.
 PR 25-OCT-1996: US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR.
 DR P-PSDB: W42087.
 PT New isolated Down's Syndrome-cell adhesion molecule - used to
 develop products for detection, diagnosis and therapy of
 developmental and neurological abnormalities
 PS Claim 3: Page 83-90: 109pp: English.
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule
 DS-CAM2 (see W42087), an extracellular, soluble protein belonging
 to a novel subclass of the Ig superfamily with highest homology to
 neural cell adhesion molecules. A modified direct cDNA selection
 technique was applied to bacterial and plasmid artificial chromosomes
 between ETS2 and MX1 by using cDNA from trisomy 21 human foetal
 brain. A unique cDNA fragment, designated E51 (see V31982), was
 obtained and used to screen a trisomy 21 human foetal brain (14-wk
 gestation) cDNA library. Further clones were obtained by exon
 trapping, and the complete DS-CAM2 cDNA sequence was identified.
 CC DS-CAM2 is a splice variant of membrane-bound DS-CAM1 (see W42086)
 and is deleted for the entire transmembrane domain. The DS-CAM
 gene spans 900-1200 kb of genomic DNA and maps at chromosome
 21q22.2-22.3. The invention also provides murine DS-CAM partial
 sequences (see V31985-87), expression vectors and host cells,
 transgenic animals, antisense oligonucleotides, and primers useful
 for amplification of DS-CAM nucleic acid. DS-CAM polypeptides are
 associated with developmental and neurological processes. They can
 be used in e.g. neural prosthetic devices used in entubulation
 methods of repairing (regenerating) damaged or severed peripheral
 nerves. The products can also be used in detection, diagnosis and
 therapy of developmental and neurological abnormalities such as
 Down syndrome, mental retardation, holoprosencephaly, agenesis of
 the corpus callosum, or schizencephaly. Antisense oligonucleotides
 are used for inhibiting translation of mRNA. 1781 C; 1707 G; 1292 T;
 SO Sequence 6413 BP; 1633 A; 1781 C; 1707 G; 1292 T;

Query Match 100.0%; Score 20; DB 1; Length 6413;
 Best Local Similarity 100.0%; Pred. No. 0.62;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccaagttcctaagagcagcag 20
 DB 4877 CCAAGTTCTCAAGAGCAGCAG 4896
 |||||||||||||||||||

RESULT 3
 ID V31981
 AC V31981 standard; cDNA: 6604 BP.
 DE 28-SEP-1998 (first entry)
 KW Human Down syndrome-cell adhesion molecule; neural cell;
 KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KW holoprosencephaly; corpus callosum agenesis;
 KW schizencephaly; diagnosis; assay; human; ds; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 453..6185
 FT /*tag= a
 FT sig_peptide 452..521
 FT /*tag= b
 FT mat_peptide 522..6162
 FT /*tag= c
 PN MO9817795-A1.
 PD 30-APR-1998.
 PF 23-OCT-1997: U19547.

PR 25-OCT-1996: US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR.
 DR WPI: 98-271791/24.
 DR P-PSDB: W42086.
 PT New isolated Down's Syndrome-cell adhesion molecule - used to
 develop products for detection, diagnosis and therapy of
 developmental and neurological abnormalities
 PS Claim 3: Page 65-72: 109pp: English.
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule
 DS-CAM1 (see W42086), a cell surface glycoprotein belonging to a
 novel subclass of the Ig superfamily with highest homology to
 neural cell adhesion molecules. A modified direct cDNA selection
 technique was applied to bacterial and plasmid artificial chromosomes
 between ETS2 and MX1 by using cDNA from trisomy 21 human foetal
 brain. A unique cDNA fragment, designated E51 (see V31982), was
 obtained and used to screen a trisomy 21 human foetal brain (14-wk
 gestation) cDNA library. Further clones were obtained by exon
 trapping. The complete DS-CAM1 cDNA sequence was deduced from
 overlapping clones. A splice variant cDNA (see V31988), encoding
 non-membrane bound DS-CAM2 (see W42087), was also identified. The
 DS-CAM gene spans 900-1200 kb of genomic DNA and maps at chromosome
 21q22.2-22.3. The invention also provides murine DS-CAM partial
 sequences (see V31985-87), expression vectors and host cells,
 transgenic animals, antisense oligonucleotides, and primers useful
 for amplification of DS-CAM nucleic acid. DS-CAM polypeptides are
 associated with developmental and neurological processes. They can
 be used in e.g. neural prosthetic devices used in entubulation
 methods of repairing (regenerating) damaged or severed peripheral
 nerves. The products can also be used in detection, diagnosis and
 therapy of developmental and neurological abnormalities such as
 Down syndrome, mental retardation, holoprosencephaly, agenesis of
 the corpus callosum, or schizencephaly. Antisense oligonucleotides
 are used for inhibiting translation of mRNA. 1673 A; 1827 C; 1768 G; 1336 T;
 SO Sequence 6604 BP; 1673 A; 1827 C; 1768 G; 1336 T;

Query Match 100.0%; Score 20; DB 1; Length 6604;
 Best Local Similarity 100.0%; Pred. No. 0.62;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccaagttcctaagagcagcag 20
 DB 4877 CCAAGTTCTCAAGAGCAGCAG 4896
 |||||||||||||||||||

RESULT 4
 ID V31987
 AC V31987 standard; cDNA: 2174 BP.
 DE 28-SEP-1998 (first entry)
 KW Mouse Down syndrome-cell adhesion molecule; DS-CAM 3' cDNA.
 KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KW holoprosencephaly; corpus callosum agenesis;
 KW schizencephaly; diagnosis; assay; mouse; ds; ss.
 OS Mus sp.
 PN MO9817795-A1.
 PD 30-APR-1998.
 PF 23-OCT-1997: U19547.
 PR 25-OCT-1996: US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR.
 DR WPI: 98-271791/24.
 PT New isolated Down's Syndrome-cell adhesion molecule - used to
 develop products for detection, diagnosis and therapy of
 developmental and neurological abnormalities
 PS Claim 2: Page 81-83: 109pp: English.
 CC This cDNA sequence comprises the 3' region of a cDNA clone
 for murine Down syndrome-cell adhesion molecule (DS-CAM), a
 member of a novel subclass of the Ig superfamily with homology to
 neural cell adhesion molecules. The middle region (see V31986)
 and 5' region (see V31985) of the clone are also provided. The

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:24:19 ; Search time 216.06 Seconds
(without alignments)
11.082 Million cell updates/sec

Title: US-08-956-991-5

Perfect score: 20
Sequence: 1 ccaggtctcaaggagcagg 20

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents.NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/5E.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/5F.COMB.seq:*
7: /cgn2_6/ptodata/1/ina/5G.COMB.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.2	76.0	3417	3	US-08-464-402-1 Sequence 1, Appl 1
2	15.2	76.0	8298	6	PCT-US93-03076-1 Sequence 1, Appl 1
3	15.2	76.0	12752	3	US-08-459-146-1 Sequence 1, Appl 1
4	15.2	76.0	12752	3	US-08-459-065-1 Sequence 1, Appl 1
5	14.4	72.0	1314	1	US-07-662-005A-15 Sequence 15, Appl 1
6	14.4	72.0	3648	2	US-08-053-614-1 Sequence 1, Appl 1
7	14.4	72.0	3648	2	US-09-316-397B-1 Sequence 1, Appl 1
8	14.4	72.0	3648	2	US-09-034-306-1 Sequence 1, Appl 1
9	14.4	72.0	3648	6	PCT-US93-09782-1 Sequence 1, Appl 1
10	14.4	72.0	4821	2	US-08-053-614-3 Sequence 3, Appl 1
11	14.4	72.0	4821	2	US-08-316-397B-3 Sequence 3, Appl 1
12	14.4	72.0	4821	2	US-09-034-306-3 Sequence 3, Appl 1
13	14.4	72.0	4821	6	PCT-US93-09782-3 Sequence 3, Appl 1
14	14.4	72.0	10299	4	US-08-477-451-1 Sequence 1, Appl 1
15	14.4	72.0	10299	4	US-08-477-451-5 Sequence 5, Appl 1
16	14.4	72.0	19932	4	US-08-477-451-25 Sequence 25, Appl 1
17	14.4	71.0	460	4	US-08-477-727A-1 Sequence 1, Appl 1
18	14.2	71.0	580	2	US-08-272-255-19 Sequence 19, Appl 1
19	14.2	71.0	580	6	PCT-US95-08565-19 Sequence 19, Appl 1
20	14.2	71.0	874	3	US-08-850-119-2 Sequence 2, Appl 1
21	14.2	71.0	911	2	US-08-745-603-1 Sequence 1, Appl 1
22	14.2	71.0	1569	4	US-08-923-772-1 Sequence 1, Appl 1
23	14.2	71.0	1617	4	US-08-324-483-1 Sequence 1, Appl 1
24	14.2	71.0	1686	1	US-08-324-483-1 Sequence 1, Appl 1
25	14.2	71.0	1747	3	US-08-522-421-5 Sequence 5, Appl 1
26	14.2	71.0	1812	4	US-08-735-041A-3 Sequence 3, Appl 1
27	14.2	71.0	2103	4	US-08-735-041A-5 Sequence 5, Appl 1

28	14.2	71.0	2103	4	US-08-735-041A-7 Sequence 7, Appl 1
29	14.2	71.0	2559	3	US-08-886-152-4 Sequence 4, Appl 1
30	14.2	71.0	2692	1	US-07-932-454A-2 Sequence 2, Appl 1
31	14.2	71.0	3171	6	PCT-US95-09261-1 Sequence 1, Appl 1
32	14.2	71.0	3215	2	US-08-426-627-1 Sequence 1, Appl 1
33	14.2	71.0	4104	1	US-07-998-003A-94 Sequence 94, Appl 1
34	14.2	71.0	4104	1	US-08-453-274B-94 Sequence 94, Appl 1
35	14.2	71.0	4104	2	US-08-453-695A-94 Sequence 94, Appl 1
36	14.2	71.0	4104	2	US-08-453-695A-94 Sequence 94, Appl 1
37	14.2	71.0	4104	3	US-08-453-702A-94 Sequence 94, Appl 1
38	14.2	71.0	4104	6	PCT-US93-12588-94 Sequence 94, Appl 1
39	14.2	71.0	4104	6	PCT-US95-08071-94 Sequence 94, Appl 1
40	14.2	71.0	4650	1	US-07-998-003A-102 Sequence 102, Appl 1
41	14.2	71.0	4650	1	US-08-453-274B-102 Sequence 102, Appl 1
42	14.2	71.0	4650	2	US-08-453-695A-102 Sequence 102, Appl 1
43	14.2	71.0	4650	2	US-08-268-161A-102 Sequence 102, Appl 1
44	14.2	71.0	4650	3	US-08-453-702A-102 Sequence 102, Appl 1
45	14.2	71.0	4650	6	PCT-US93-12588-102 Sequence 102, Appl 1

ALIGNMENTS

RESULT 1
US-08-464-402-1
Sequence 1, Application US/08464402
Patent No. 5858705
GENERAL INFORMATION:
APPLICANT: WEI, ET AL.
TITLE OF INVENTION: Human DNA Ligase III
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INC DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,402
FILING DATE: June 5, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/03939
FILING DATE: 31 MAR 95
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3417 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-464-402-1
Query Match 76.0%; Score 15.2; DB 3; Length 3417;
Best Local Similarity 85.0%; Pred. NO. 41;
Matches 17; Conservative 0; Mismatches 3; Indels 0;
Gaps 0;
G 1 ccaggtctcaaggagcagg 20

||||| ||| ||| |||
Db 2559 CCACTTCTCCAGCAGCAGC 2578

RESULT 2
PCT-US93-03076-1
; Sequence 1, Application PC/TUS9303076
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; TITLE OF INVENTION: GAP-Associated Protein p190 and
; TITLE OF INVENTION: Transduction
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03076
; FILING DATE: 19930331
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH192-03A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8298 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 731..5272
; PCT-US93-03076-1

Query Match 76.0%; Score 15.2; DB 6; Length 8298;
Best Local Similarity 85.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccagttctcaagcagcagc 20
Db 3763 CAAATTCTCAATGAGCTGG 3782

RESULT 3
US-08-459-146-1
; Sequence 1, Application US/08459146
; Patent No. 5886405
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.

ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,146
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 74,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12752 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryptomycetozoa)
; ORGANISM: parasitica
; STRAIN: EP713
; US-08-459-146-1

Query Match 76.0%; Score 15.2; DB 3; Length 12752;
Best Local Similarity 85.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccagttctcaagcagcagc 20
Db 12192 CCAACTCGCAAGCAGCAGC 12211

RESULT 4
US-08-459-065-1
; Sequence 1, Application US/08459065
; Patent No. 5882642
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,065
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,117

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:36 ; Search time 4987.54 Seconds
(without alignments)
15.140 Million cell updates/sec

Title: US-08-956-991-5
Perfect score: 20
Sequence: 1 ccagttccaaagagagcagc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: gb_est33:*
53: gb_est34:*
54: gb_est35:*
55: gb_est36:*
56: gb_est37:*
57: gb_est38:*
58: gb_est39:*
59: gb_est40:*
60: gb_est41:*
61: gb_est42:*
62: gb_est43:*
63: gb_est44:*
64: gb_est45:*
65: gb_est46:*
66: gb_est47:*
67: gb_est48:*
68: gb_est49:*
69: gb_est50:*
70: gb_est51:*
71: gb_est52:*
72: gb_est53:*
73: gb_est54:*
74: gb_est55:*
75: gb_est56:*
76: gb_est57:*
77: gb_est58:*
78: gb_est59:*
79: gb_est60:*
80: gb_est61:*
81: gb_est62:*
82: gb_est63:*
83: gb_est64:*
84: gb_est65:*
85: gb_est66:*
86: gb_est67:*
87: gb_est68:*
88: gb_est69:*
89: gb_est70:*
90: gb_est71:*
91: gb_est72:*
92: gb_est73:*
93: gb_est74:*
94: gb_est75:*
95: gb_est76:*
96: gb_est77:*
97: gb_est78:*
98: gb_est79:*
99: gb_est80:*
100: gb_est81:*
101: gb_est82:*
102: gb_est83:*
103: gb_est84:*
104: gb_est85:*
105: gb_est86:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.8	84.0	177	74	AM176762
2	16.8	84.0	319	61	AV158019

```

3 16.8 84.0 334 46 A1422998 tcttbb11.x
4 16.8 84.0 343 62 A1927023 w08tA06.x
5 16.8 84.0 356 24 N24517 N24917 yy01h12.s1
6 16.8 84.0 358 63 AM015308 AM015308 UI-H-B10-
7 16.8 84.0 359 39 C84007 C84007 dict
8 16.8 84.0 360 45 A1347190 A1347190 tcd04g04.x
9 16.8 84.0 363 30 AA200841 AA200841 mu03g06.t
10 16.8 84.0 400 88 AO894484 HS_2036.B
11 16.8 84.0 449 29 AA188488 zq44d06.s
12 16.8 84.0 469 49 A1636091 t292f09.x
13 16.8 84.0 482 91 AO150756 HS_3091.A
14 16.8 84.0 526 29 AA1607977 AA1607977 z050e04.s
15 16.8 84.0 583 100 AO307774 HS_2175_A
16 16.8 84.0 725 45 AU005349 AU005349 AU005349
17 16.8 84.0 856 32 AO739885 HS_5505.A
18 16.8 84.0 224 37 AA694519 ah35h05.s
19 16.4 82.0 312 100 AO339036 AO339036 HS_5021_A
20 16.4 82.0 390 38 AA770324 ah83a04.s
21 16.4 82.0 483 26 W58253 W58253 z025d02.t1
22 16.4 82.0 487 40 AA914803 AA914803 v209d01.r
23 16.4 82.0 488 24 H94418 H94418 yv18d12.t1
24 16.4 82.0 550 103 AO479175 AO479175 RPEC1-11-2
25 16.4 82.0 569 30 AA268152 va40b07.r
26 16.4 82.0 572 39 AA906943 cV95b07.s
27 16 80.0 410 102 AO356983 AO356983 CTBTB1-E1-
28 16 80.0 428 91 AO147488 AO147488 HS_3070_A
29 16 80.0 542 102 AO357152 AO357152 CTBTB1-E1-
30 15.8 79.0 174 105 AO332354 AO332354 RPEC1-11-4
31 15.8 79.0 186 47 A1500844 A1500844 UI-R-C2P-
32 15.8 79.0 206 69 AV218986 AV218986 AV218986
33 15.8 79.0 272 31 AA319713 EST21997 AA319713 EST21997
34 15.8 79.0 316 59 AV102330 AV102330 AV102330
35 15.8 79.0 332 27 AA018975 AA018975 z055e09.r
36 15.8 79.0 340 39 AA893694 AA893694 EST197497
37 15.8 79.0 344 47 A1508909 A1508909 VC46c04.Y
38 15.8 79.0 370 48 A1598398 A1598398 EST250101
39 15.8 79.0 370 48 A1598669 A1598669 EST250372
40 15.8 79.0 382 46 AA963736 AA963736 UI-R-C0-9
41 15.8 79.0 403 49 C96311 C96311 C96311
42 15.8 79.0 406 47 A1502552 A1502552 UI-R-C2-n
43 15.8 79.0 409 46 AA99106 UI-R-C0-1
44 15.8 79.0 415 21 R02306 R02306 y082a01.s1
45 15.8 79.0 416 41 A1005674 A1005674 CV59f07.s

```

ALIGNMENTS

```

RESULT 1
LOCUS AM176762 177 bp mRNA EST 16-NOV-1999
DEFINITION RCO-CT0094-200899-002-B12 CT0094 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM176762
VERSION AM176762.1 GI:6442799
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 177)
AUTHORS HCGP h"p://www.ludwig.org.br/ORESTES".
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138231.

```

```

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome

```

```

FEATURES
SOURCE
location/Qualifiers
1..177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0094"
/dev_stage="Adult"
/Note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (O.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT
45 a 38 c 62 g 32 t
ORIGIN

```

```

Query Match 84.0%; Score 16.8; DB 74; Length 177;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ccaggtctcaaggagcagg 20
DB 37 CCAGGTGTCGAAGGACAGC 56

```

```

RESULT 2
LOCUS AV1 8019 319 bp mRNA EST 07-JUL-1999
DEFINITION AV1_3019 Mus musculus head C57BL/6J 12-day embryo Mus musculus cDNA
ACCESSION AV18019
VERSION AV18019.1 GI:5401654
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 319)
AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Mitsuuma, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraishi, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Riken Mouse ESTs
Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3137650.
CONTACT: Chile Owa
Genome Science Laboratory
RIKEN

```

```

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rcc.riken.go.jp
Thermostabilization and thermoinactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rcc.riken.go.jp) for
further details.

```

```

FEATURES
SOURCE
1..319
/organism="Mus musculus"
/strain="C57BL/6J"

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:31:15 : Search time 383.18 Seconds
(without alignments)
13.059 Million cell updates/sec

Title: US-08-956-991-6

Perfect score: 20
Sequence: 1 cctgtatgacctgcaggaag 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	V31984	Human Down syndrome
2	20	100.0	388	V31982	Human Down syndrome
3	20	100.0	6604	V31981	Human Down syndrome
4	18.4	92.0	6413	V31988	Human Down syndrome
5	17.4	87.0	2173	V31987	Mouse Down syndrome
6	16.4	82.0	847	V59703	Human secreted pro
7	16.4	82.0	852	V59704	Human secreted pro
8	16.4	82.0	1250	X24827	Calcium activated
9	16.4	82.0	1287	V35463	Human hiki coding
10	16.4	82.0	1354	V59705	Human secreted pro
11	16.4	82.0	1486	V59523	Human secreted pro
12	16.4	82.0	1524	X24828	Calcium activated
13	16.4	82.0	2238	X24826	Calcium activated
14	16.4	82.0	2261	X24825	Human calcium acti
15	15.8	79.0	1285	Q20229	Sequence of the mu
16	15.8	79.0	1296	Q20228	ADP- glucose-pyroph
17	15.8	79.0	1296	Q80082	ADP- glucose-pyroph
18	15.8	79.0	1296	Q80083	ADP- glucose-pyroph
19	15.8	79.0	1322	Q25979	gluc. glycogen bio
20	15.8	79.0	1322	V23650	E. coli ADP- glucos
21	15.8	79.0	3955	T01590	Rat neu promoter.
22	15.4	77.0	24	Q45927	GCN4-2, GCN4 bindi
23	15.4	77.0	24	Q45929	GCN4-4, GCN4 bindi
24	15.4	77.0	5102	V10261	Human COL4A6 gene.
25	15.2	76.0	22	Q67130	CD40 ligand gene m
26	15.2	76.0	50	X52145	Synthetic plasmid
27	15.2	76.0	96	X03315	Test RNA fragment
28	15.2	76.0	717	V69319	Neomycin resistanc
29	15.2	76.0	818	V13141	np1 coding regio
30	15.2	76.0	1003	V13140	np11 CDS fused to
31	15.2	76.0	1186	Q25707	Chimeric neo gene
32	15.2	76.0	1321	V44078	mu-NDO-r DNA fragm
33	15.2	76.0	1415	Q28574	Sequence of the do
34	15.2	76.0	1521	V14348	Plasmid pCMVgfo11

35	15.2	76.0	1521	V14353	Plasmid pPKgfo25
36	15.2	76.0	1851	T90372	Recombinant molecu
37	15.2	76.0	1872	Q06828	Extracellular port
38	15.2	76.0	2118	T90371	Recombinant molecu
39	15.2	76.0	2224	T90373	Recombinant molecu
40	15.2	76.0	2239	X08974	RelA-bacil encodin
41	15.2	76.0	2385	T72725	Her2-GM-CSF immuno
42	15.2	76.0	2447	Q30000	Chimeric lux::npt
43	15.2	76.0	2526	T62395	Construct pMK10/AS
44	15.2	76.0	2529	Q50489	Promoterless neo/a
45	15.2	76.0	2549	Q56107	DNA encoding a 11p

ALIGNMENTS

RESULT 1	
V31984	
ID V31984 standard; cDNA; 20 BP.	
AC V31984:	
DT 28-SEP-1998 (first entry)	
DE Human Down syndrome-cell adhesion molecule DS-CAM primer B9-131R.	
KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; PCR; primer; ss.	
OS Synthetic.	
OS Homo sapiens.	
PN W09817795-A1.	
PD 30-APR-1998.	
PF 23-OCT-1997; U19547.	
PR 25-OCT-1996; US-029322.	
PA (CEDA-) CEDARS SINAI MEDICAL CENT.	
PI Kolenberg, JR.	
DR WPI: 98-271791/24.	
PT New isolated Down's syndrome-cell adhesion molecule - used to	
PT develop products for detection, diagnosis and therapy of	
PT developmental and neurological abnormalities	
PS Example 5: Page 79: 109pp; English.	
CC Reverse primer B9-131R and forward primer B9-131F (see V31983)	
CC were used in RT-PCR assays of cDNA libraries of various human	
CC tissues to determine human Down syndrome-cell adhesion molecule	
CC DS-CAM (see V31981 and V31988) expression. The results	
CC demonstrated expression of human DS-CAM mRNA in foetal and adult	
CC brain, and foetal kidney. In addition, a breast carcinoma cell	
CC line showed expression of human DS-CAM mRNA. DS-CAM polypeptides	
CC (see W42086-87) are associated with developmental and neurological	
CC processes. The polypeptides and nucleic acids are used to develop	
CC products for the detection, diagnosis and therapy of developmental	
CC and neurological abnormalities.	
SQ Sequence 20 BP: 5 A; 5 C; 6 G; 4 T;	
Query Match 100.0%; Score 20; DB 1; Length 20;	
Best Local Similarity 100.0%; Pred. No. 0.77;	
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 cctgtatgacctgcaggaag 20	
DB 1 CCIGTATGACCTGCAGGAAG 20	
RESULT 2	
V31982/c	
ID V31982 standard; cDNA; 388 BP.	
AC V31982:	
DT 28-SEP-1998 (first entry)	
DE Human Down syndrome-cell adhesion molecule DS-CAM probe E51.	
KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; ds; ss.	
OS Homo sapiens.	

PN W09817795-A1.
 PD 30-APR-1998.
 PF 23-OCT-1997: U19547.
 PR 25-OCT-1996: US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR:
 DR WPI: 98-271791/24.
 PT New isolated Down's Syndrome-cell adhesion molecule - used to develop products for detection, diagnosis and therapy of developmental and neurological abnormalities
 PS Example 2: Page 79: 109pp: English.
 CC This polynucleotide comprises cDNA fragment E51 that was isolated from a human trisomy 21 foetal brain (14-wk gestation) cDNA library following a modified direct cDNA selection technique applied to bacterial and pl artificial chromosomes between ERS2 and MX1. E51 was used as a probe to screen the trisomy 21 foetal brain library. 62 clones were isolated from the 2 million clones of the original library. Overlapping clones were obtained that were used to deduce a full-length sequence (see V31981) coding for novel Down syndrome-cell adhesion molecule DS-CAM1 (see M42086). A splice variant, DS-CAM2 (see V31919), was also identified. DS-CAM polypeptides are associated with developmental and neurological processes. The polypeptides and nucleic acids are used to develop products for the detection, diagnosis and therapy of developmental and neurological abnormalities.
 SO Sequence 388 BP: 83 A: 119 C: 95 G: 91 T:

Query Match 100.0%; Score 20; DB 1; Length 388;
 Best Local Similarity 100.0%; Pred. No. 0.94;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctgtatgactgcaggag 20
 Db 331 CCTGTATGACTGCAGGAG 312

RESULT 3
 V31981
 ID V31981 standard: cDNA: 6604 BP.
 AC V31981;
 DT 28-SEP-1998 (first entry)
 DE Human Down syndrome-cell adhesion molecule DS-CAM1 cDNA.
 KM DS-CAM1: Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KM holoprosencephaly; corpus callosum agenesis;
 KM schizencephaly; diagnosis: assay; human; ds; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 453..6185
 FT sig_peptide 452..521
 FT mat_peptide 522..6162
 FT /tag= b
 FT /tag= D
 FT /tag= c
 PN W09817795-A1.
 PD 30-APR-1998.
 PF 23-OCT-1997: U19547.
 PR 25-OCT-1996: US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR:
 DR WPI: 98-271791/24.
 P-PSDB: M42086.
 PT New isolated Down's Syndrome-cell adhesion molecule - used to develop products for detection, diagnosis and therapy of developmental and neurological abnormalities
 PS Claim 3: Page 65-72: 109pp: English.
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule DS-CAM1 (see M42086), a cell surface glycoprotein belonging to a novel subclass of the Ig superfamily with highest homology to neural cell adhesion molecules. A modified direct cDNA selection technique was applied to bacterial and pl artificial chromosomes between ERS2 and MX1 by using cDNA from trisomy 21 human foetal

CC brain. A unique cDNA fragment, designated E51 (see V31982), was obtained and used to screen a trisomy 21 human foetal brain (14-wk gestation) cDNA library. Further clones were obtained by exon trapping. The complete DS-CAM1 cDNA sequence was deduced from overlapping clones. A splice variant cDNA (see V31988), encoding non-membrane bound DS-CAM2 (see M42087), was also identified. The CC DS-CAM gene spans 900-1200 kb of genomic DNA and maps at chromosome 21q22.2-22.3. The invention also provides murine DS-CAM partial sequences (see V31985-87), expression vectors and host cells, transgenic animals, antisense oligonucleotides, and primers useful CC for amplification of DS-CAM nucleic acid. DS-CAM polypeptides are CC associated with developmental and neurological processes. They can CC be used in e.g. neural prosthetic devices used in entubulation CC methods of repairing (regenerating) damaged or severed peripheral CC nerves. The products can also be used in detection, diagnosis and CC therapy of developmental and neurological abnormalities such as CC Down syndrome, mental retardation, holoprosencephaly, agenesis of CC the corpus callosum, or schizencephaly. Antisense oligonucleotides CC are used for inhibiting translation of mRNA.
 SO Sequence 6604 BP: 1673 A: 1827 C: 1768 G: 1336 T:

Query Match 100.0%; Score 20; DB 1; Length 6604;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctgtatgactgcaggag 20
 Db 5054 CCTGTATGACTGCAGGAG 5073

RESULT 4
 V31988
 ID V31988 standard: cDNA: 6413 BP.
 AC V31988;
 DT 28-SEP-1998 (first entry)
 DE Human Down syndrome-cell adhesion molecule DS-CAM2 cDNA.
 KM DS-CAM2: Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KM holoprosencephaly; corpus callosum agenesis;
 KM schizencephaly; diagnosis: assay; human; ds; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 453..5168
 FT /tag= a
 PN W09817795-A1.
 PD 30-APR-1998.
 PF 23-OCT-1997: U19547.
 PR 25-OCT-1996: US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR:
 DR WPI: 98-271791/24.
 P-PSDB: M42087.
 PT New isolated Down's Syndrome-cell adhesion molecule - used to develop products for detection, diagnosis and therapy of developmental and neurological abnormalities
 PS Claim 3: Page 83-90: 109pp: English.
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule DS-CAM2 (see M42087), an extracellular, soluble protein belonging to a novel subclass of the Ig superfamily with highest homology to neural cell adhesion molecules. A modified direct cDNA selection technique was applied to bacterial and pl artificial chromosomes between ERS2 and MX1 by using cDNA from trisomy 21 human foetal brain. A unique cDNA fragment, designated E51 (see V31982), was obtained and used to screen a trisomy 21 human foetal brain (14-wk gestation) cDNA library. Further clones were obtained by exon trapping, and the complete DS-CAM2 cDNA sequence was identified. DS-CAM2 is a splice variant of membrane-bound DS-CAM1 (see M42086) and is deleted for the entire transmembrane domain. The DS-CAM CC gene spans 900-1200 kb of genomic DNA and maps at chromosome 21q22.2-22.3. The invention also provides murine DS-CAM partial CC sequences (see V31985-87), expression vectors and host cells, transgenic animals, antisense oligonucleotides, and primers useful

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:24:22 ; Search time 216.06 Seconds
(Without alignments)
11.082, 11111111 cell updates/sec

Title: US-08-956-991-6
Perfect score: 20
Sequence: 1 cctgtatgacctgcaggaag 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTUS9.COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	15.8	79.0	1296 1 US-08-090-523-1	Sequence 1, Appl1
2	15.8	79.0	1296 1 US-08-090-523-3	Sequence 3, Appl1
3	15.8	79.0	1296 1 US-08-398-627-1	Sequence 1, Appl1
4	15.8	79.0	1296 1 US-08-398-627-3	Sequence 3, Appl1
5	15.8	79.0	1296 1 US-08-406-858-1	Sequence 1, Appl1
6	15.8	79.0	1296 1 US-08-406-858-3	Sequence 3, Appl1
7	15.8	79.0	1296 6 PCT-US91-04036-1	Sequence 1, Appl1
8	15.8	79.0	1296 6 PCT-US91-04036-3	Sequence 3, Appl1
9	15.8	79.0	1296 6 PCT-US94-05275-1	Sequence 1, Appl1
10	15.8	79.0	1296 6 PCT-US94-05275-3	Sequence 3, Appl1
11	15.8	79.0	1323 1 US-07-735-065-3	Sequence 3, Appl1
12	15.8	79.0	1323 2 US-08-469-202-13	Sequence 13, Appl1
13	15.8	79.0	1323 4 US-08-484-434C-13	Sequence 13, Appl1
14	15.8	79.0	3955 1 US-08-229-515A-14	Sequence 14, Appl1
15	15.8	79.0	3955 1 US-08-645-865-14	Sequence 1, Appl1
16	15.4	77.0	24 6 PCT-US93-00321-2	Sequence 2, Appl1
17	15.4	77.0	24 6 PCT-US93-00321-4	Sequence 4, Appl1
18	15.4	77.0	5102 2 US-08-184-422-15	Sequence 15, Appl1
19	15.2	76.0	22 1 US-08-064-121-2	Sequence 2, Appl1
20	15.2	76.0	1186 1 US-08-478-015-2	Sequence 2, Appl1
21	15.2	76.0	1186 2 US-08-478-015-2	Sequence 2, Appl1
22	15.2	76.0	1186 5 US-08-478-015-2	Sequence 2, Appl1
23	15.2	76.0	1415 1 US-08-257-975-2	Sequence 8, Appl1
24	15.2	76.0	1415 1 US-08-257-975-2	Sequence 8, Appl1
25	15.2	76.0	2119 5 US-09-032-372-6	Sequence 6, Appl1
26	15.2	76.0	2120 1 US-08-404-732A-4	Sequence 4, Appl1
27	15.2	76.0	2220 3 US-08-864-224-1	Sequence 1, Appl1

28	15.2	76.0	2224 1 US-08-404-732A-8	Sequence 8, Appl1
29	15.2	76.0	2285 4 US-09-146-283-3	Sequence 3, Appl1
30	15.2	76.0	2445 1 US-08-122-520C-8	Sequence 8, Appl1
31	15.2	76.0	3176 7 5212080-1	Patent No. 5212080
32	15.2	76.0	3200 1 US-08-453-104-23	Sequence 23, Appl1
33	15.2	76.0	3200 3 US-08-694-824-23	Sequence 23, Appl1
34	15.2	76.0	3201 1 US-08-453-104-22	Sequence 22, Appl1
35	15.2	76.0	3201 3 US-08-694-824-22	Sequence 22, Appl1
36	15.2	76.0	3740 1 US-07-794-400-14	Sequence 14, Appl1
37	15.2	76.0	3740 1 US-08-041-648-3	Sequence 3, Appl1
38	15.2	76.0	3740 1 US-08-217-529-3	Sequence 3, Appl1
39	15.2	76.0	3740 1 US-08-397-470-14	Sequence 14, Appl1
40	15.2	76.0	3768 3 US-08-625-101-1	Sequence 1, Appl1
41	15.2	76.0	3768 3 US-08-356-786-1	Sequence 1, Appl1
42	15.2	76.0	3974 5 US-09-026-343-33	Sequence 33, Appl1
43	15.2	76.0	4473 4 US-09-048-804-1	Sequence 1, Appl1
44	15.2	76.0	4530 1 US-08-229-515A-9	Sequence 9, Appl1
45	15.2	76.0	4530 1 US-08-645-865-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-08-090-523-1
; Sequence 1, Application US/06090523
; Patent No. 5498830
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Enhanced Starch Biosynthesis
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Grace L. Bonner, Monsanto Co. B94F
; STREET: 700 Chesterfield Parkway No. 5498830th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/090,523
; FILING DATE: 19930712
; CLASSIFICATION: 800
; PRIOR APPLICATION NUMBER: US 07/709663
; FILING DATE: 07-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539763
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.
; REGISTRATION NUMBER: 32,963
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 537-7286
; TELEFAX: (314) 537-6047
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1293
; US-08-090-523-1

Query Match 79.0%: Score 15.8; DB 1: Length 1296;
Best Local Similarity 89.5%: Pred. No. 27;
Matches 17: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 2 ctgtatgacctgcaggaag 20
||||| ||||| |||||
Db 667 CTGTATGAAGCTGTGGAAG 685

RESULT 2

US-08-090-523-3
Sequence 3, Application US/08090523
Patent No. 5498830
GENERAL INFORMATION:
APPLICANT: Barry, Gerard F.
APPLICANT: Kishore, Ganesh M.
APPLICANT: Stark, David M.
TITLE OF INVENTION: Enhanced Starch Biosynthesis
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace L. Bonner, Monsanto Co. B94F
STREET: 700 Chesterfield Parkway No. 5498830th
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/090.523
FILING DATE: 19930712
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/709663
FILING DATE: 07-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539763
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Bonner, Grace L.
REGISTRATION NUMBER: 32,963
REFERENCE/DOCKET NUMBER: 38-21(10559)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 537-7286
TELEFAX: (314) 537-6047
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1293
US-08-090-523-3

Query Match 79.0%: Score 15.8; DB 1: Length 1296;
Best Local Similarity 89.5%: Pred. No. 27;
Matches 17: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 2 ctgtatgacctgcaggaag 20
||||| ||||| |||||
Db 667 CTGTATGAAGCTGTGGAAG 685

RESULT 3

US-08-398-627-1
Sequence 1, Application US/08398627

Patent No. 5608149
GENERAL INFORMATION:
APPLICANT: Barry, Gerard F.
APPLICANT: Kishore, Ganesh M.
APPLICANT: Stark, David M.
TITLE OF INVENTION: Enhanced Starch Biosynthesis
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace L. Bonner, Monsanto Co. B94F
STREET: 700 Chesterfield Parkway No. 5608149th
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398.627
FILING DATE: 03-MAR-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/090.523
FILING DATE: 12-JUL-1993
APPLICATION NUMBER: US 07/709663
FILING DATE: 07-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539763
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Bonner, Grace L.
REGISTRATION NUMBER: 32,963
REFERENCE/DOCKET NUMBER: 38-21(10559)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 537-7286
TELEFAX: (314) 537-6047
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1293
US-08-398-627-1

Query Match 79.0%: Score 15.8; DB 1: Length 1296;
Best Local Similarity 89.5%: Pred. No. 27;
Matches 17: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 2 ctgtatgacctgcaggaag 20
||||| ||||| |||||
Db 667 CTGTATGAAGCTGTGGAAG 685

RESULT 4

US-08-398-627-3
Sequence 3, Application US/08398627
Patent No. 5608149
GENERAL INFORMATION:
APPLICANT: Barry, Gerard F.
APPLICANT: Kishore, Ganesh M.
APPLICANT: Stark, David M.
TITLE OF INVENTION: Enhanced Starch Biosynthesis
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:40 ; Search time 4987.54 Seconds
(without alignments)
15.140 Million cell updates/sec

Title: US-08-956-991-6
Perfect score: 20
Sequence: 1 ccgtatgacctgcaggaag 20

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 5077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: qb_est1:*
21: qb_est2:*
22: qb_est3:*
23: qb_est4:*
24: qb_est5:*
25: qb_est6:*
26: qb_est7:*
27: qb_est8:*
28: qb_est9:*
29: qb_est10:*
30: qb_est11:*
31: qb_est12:*
32: qb_est13:*
33: qb_est14:*
34: qb_est15:*
35: qb_est16:*
36: qb_est17:*
37: qb_est18:*
38: qb_est19:*
39: qb_est20:*
40: qb_est21:*
41: qb_est22:*
42: qb_est23:*
43: qb_est24:*
44: qb_est25:*

45: qb_est26:*
46: qb_est27:*
47: qb_est28:*
48: qb_est29:*
49: qb_est30:*
50: qb_est31:*
51: qb_est32:*
52: em_est0:*
53: em_est1:*
54: em_est2:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: qb_est33:*
60: qb_est34:*
61: qb_est35:*
62: qb_est36:*
63: qb_est37:*
64: qb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: qb_est39:*
70: qb_est40:*
71: qb_est41:*
72: qb_est42:*
73: qb_est43:*
74: qb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: qb_gss1:*
80: qb_gss2:*
81: qb_gss3:*
82: qb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: qb_gss5:*
88: qb_gss6:*
89: qb_gss7:*
90: qb_gss8:*
91: qb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: qb_gss10:*
100: qb_gss11:*
101: em_gss12:*
102: qb_gss12:*
103: qb_gss13:*
104: qb_gss14:*
105: qb_gss15:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	188	50	F25472
2	16.8	84.0	334	61	A1841727

```

C 3 16.8 84.0 338 21 F00929
4 16.8 84.0 339 39 AA855923
5 16.8 84.0 349 61 AA855923
6 16.8 84.0 349 61 AA855923
7 16.8 84.0 350 62 AA855923
8 16.8 84.0 415 36 AA656681
9 16.8 84.0 485 87 AA815436
10 16.8 84.0 516 34 AA498668
11 16.8 84.0 522 30 AA204382
12 16.8 84.0 550 42 AA116267
13 16.8 84.0 571 103 AA038215
14 16.8 84.0 593 63 AA1988028
15 16.8 84.0 600 62 AA1905935
16 16.8 84.0 649 74 AA177933
17 16.8 84.0 707 35 AA545906
18 16.8 84.0 734 60 AA1788686
19 16.8 84.0 745 40 AA980418
20 16.8 84.0 788 60 AA1788722
21 16.4 82.0 253 28 AA066808
22 16.4 82.0 260 90 AA058631
23 16.4 82.0 410 28 AA076337
24 16.4 82.0 479 74 AA201060
25 16.4 82.0 509 74 AA200065
26 16.4 82.0 515 105 AA0600786
27 16.4 82.0 704 59 AA1768257
28 16.4 82.0 713 88 AA0853594
29 16.4 82.0 728 82 AA0889443
30 16.4 80.0 575 63 AA1997586
31 15.8 79.0 133 26 AA59978
32 15.8 79.0 166 30 AA247199
33 15.8 79.0 180 22 AA58397
34 15.8 79.0 186 22 AA58397
35 15.8 79.0 200 59 AA1769721
36 15.8 79.0 217 25 AA10919
37 15.8 79.0 235 25 AA87427
38 15.8 79.0 242 22 AA57457
39 15.8 79.0 248 30 AA216055
40 15.8 79.0 255 25 AA84658
41 15.8 79.0 257 31 AA328566
42 15.8 79.0 259 25 AA85521
43 15.8 79.0 261 33 AA382414
44 15.8 79.0 267 29 AA196406
45 15.8 79.0 272 27 AA05020

```

ALIGNMENTS

```

RESULT 1
F25472 188 bp mRNA EST 13-MAY-1999
DEFINITION HSD12539 HM3 Homo sapiens cDNA clone s4000038E11, mRNA sequence.
ACCESSION F25472.1 GI:4811098
VERSION F25472.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 188)
AUTHORS Lanfranchi,G., Muraro,T., Caldara,F., Pacchioni,B., Pallavicini,A.,
Pandojlo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.
TITLE Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
JOURNAL Genome Res. 6 (1), 35-42 (1996)
MEDLINE 96276048
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246745.
CONTACT: Valle G.
CIRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at

```

http://grup.bio.unipd.it.
Location/Qualifiers

FEATURES

source

```

1. 188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s4000038E11"
/clone_lib="HM3"
/sex="female"
/tissue-type="pectoral muscle (after mastectomy)"
/note="Vector: pCDNAII (Invitrogen); Site_1: BstXI;
Site_2: NotI. The library is not subtracted nor normalized.
Lanfranchi. This library is not subtracted nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCGGCTCGAGCGCGCTTTTCTTTTCTTTT-3'). The
ds cDNA was sonicated and size-selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adapters, NotI digested and
directionally cloned into BstXI-NotI cut pCDNAII vector."
BASE COUNT 67 a 44 c 49 g 28 t
ORIGIN

```

Query Match 84.0%; Score 16.8; DB 50; Length 188;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Caps 0;

Qy 1 cctgtatgacctgcaggaag 20
Db 73 cctgtatgacctgcaggaag 92

```

RESULT 2
A1841727 334 bp mRNA EST 14-JUL-1999
DEFINITION UT-M-AL0-abo-f-04-0-UI-s1 NIH-BMAP_MCO Mus musculus cDNA clone
ACCESSION A1841727
VERSION A1841727.1 GI:5475940
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 334)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On May 9, 1996 this sequence version replaced gi:1132832.
CONTACT: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized prefrontal cortex library cDNA library preparation:
M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be
made available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA-Tes.

```

FEATURES
source Location/Qualifiers
1. 334

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:38:52 ; Search time 7810.03 Seconds
(without alignments)
-327.353 Million cell updates/sec

Title: US-08-956-991-7

Perfect score: 842
Sequence: 1 ccggcgccggcgccggcgag.....cgcgtgagcctctgcga 842

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_da1.*
2: gb_da2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pl4.*
11: gb_pl5.*
12: gb_pl6.*
13: gb_pl7.*
14: gb_pl8.*
15: gb_pl9.*
16: gb_pl10.*
17: gb_pl11.*
18: gb_pl12.*
19: gb_pl13.*
20: gb_pl14.*
21: gb_pl15.*
22: gb_pl16.*
23: gb_pl17.*
24: gb_pl18.*
25: gb_pl19.*
26: gb_pl20.*
27: gb_pl21.*
28: gb_pl22.*
29: gb_pl23.*
30: gb_pl24.*
31: gb_pl25.*
32: gb_pl26.*
33: gb_pl27.*
34: gb_pl28.*
35: gb_pl29.*
36: gb_pl30.*
37: gb_pl31.*
38: gb_pl32.*
39: gb_pl33.*
40: gb_pl34.*
41: gb_pl35.*
42: gb_pl36.*
43: gb_pl37.*
44: gb_pl38.*

45: gb_hlg7.*
46: em_hlg7.*
47: em_hlg8.*
48: em_hlg9.*
49: em_hlg10.*
50: gb_pl13.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	670.2	79.6	6413	11	AF023450	AF023450 Homo sapi
2	594	70.5	6110	11	AF023449	AF023449 Homo sapi
3	292.2	34.7	119678	11	AF064866	AF064866 Homo sapi
4	292.2	34.7	120007	11	AF064864	AF064864 Homo sapi
5	148.2	17.6	189607	32	AP000757	AP000757 Homo sapi
6	148.2	17.6	223724	32	AP000834	AP000834 Homo sapi
7	125.6	14.9	155407	11	AF042090	AF042090 Homo sapi
8	77.6	9.2	109866	11	AF043945	AF043945 Homo sapi
9	58.8	7.0	124395	44	AC013397	AC013397 Homo sapi
10	51.4	6.1	66362	43	AC013815	AC013815 Homo sapi
11	49	5.8	79333	35	AC005452	AC005452 Drosophila
12	46.2	5.5	69128	45	AC017764	AC017764 Drosophila
13	41.8	5.0	107475	45	AC017106	AC017106 Homo sapi
14	41	4.9	1970	4	GCY17794	Y17794 Gallus gall
15	41	4.9	162648	43	AC010827	AC010827 Homo sapi
16	40.4	4.8	417	34	AB017336	AB017336 Leishman
17	40	4.8	80908	11	HS524E15	284485 Human DNA s
18	39.8	4.7	308292	45	AC012384	AC012384 Homo sapi
19	39.4	4.7	73511	11	HS6546	292546 Human DNA s
20	39.4	4.7	1237	9	AP000353	AP000353 Homo sapi
21	39.2	4.7	66	16	HEH5Y1G3	X06461 Herpes simp
22	39.2	4.7	7660	40	AF116458	AF116458 Homo sapi
23	39.2	4.7	12001	5	AR048721	AR048721 Sequence
24	39.2	4.7	26245	16	HS1US	L00036 HSV1 (strat
25	39.2	4.7	26245	16	HS1US	L00036 HSV1 (strat
26	39.2	4.7	43034	11	HSN5H6	293024 Human DNA s
27	39.2	4.7	152261	16	HE1CC	X14112 Herpes simp
28	39.2	4.7	152261	16	HE1CC	X14112 Herpes simp
29	38.6	4.6	5887	9	SS1132828	X132828 Spermatoz
30	38.6	4.6	5887	9	HEPNC-1	X58709 Human type
31	38.6	4.6	177241	11	HS402G11	AL023238 Human DNA
32	38.4	4.6	1959	5	IO1837	IO1837 Sequence 4
33	38.4	4.6	3573	5	SCINHRAR	X03266 Porcine mRN
34	38.4	4.6	3588	5	IO52233	IO52233 Sequence 15
35	38.4	4.6	3588	5	IO52233	IO52233 Sequence 15
36	38.4	4.6	3588	5	IO52233	IO52233 Sequence 15
37	38.4	4.6	3588	5	IO52233	IO52233 Sequence 15
38	38.2	4.5	1525	4	CHKESTRP13	D26328 Gallus gall
39	38.2	4.5	7618	4	XB28SRNA	X59733 X borealis
40	38.2	4.5	81463	45	AC016882	AC016882 Homo sapi
41	38	4.5	1478	1	SC110F3	X77040 S. coelicolo
42	38	4.5	1524	5	AR065197	AR065197 Sequence
43	38	4.5	2900	1	PG1IP	X70354 P. glumae ge
44	38	4.5	2900	5	AL6323	AL6323 lIPA gene.
45	38	4.5	2900	5	AR038049	AR038049 Sequence

ALIGNMENTS

RESULT 1
AF023450 AF023450 6413 bp mRNA
LOCUS Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (USCAM)
DEFINITION mRNA, complete cds.
ACCESSION AF023450
VERSION AF023450.1 GI:3169767

100.0	842	1	V31985
79.6	6413	1	V31988
79.6	6604	1	V31981
36.5	1493	1	V27207
4.7	4257	1	V10362
4.7	4257	1	V68520
4.7	12001	1	O76213
4.6	3588	1	N70317
4.6	2900	1	O20372
4.3	3694	1	O36930
4.3	1524	1	T63571
4.1	2312	1	X16152
4.1	3608	1	N91777
4.1	3802	1	O87979
4.1	3802	1	T70227
4.1	3802	1	T66812
4.1	50341	1	V22674
4.1	52297	1	T51411
4.1	52298	1	O47357
4.1	1910	1	V48148
4.1	1170	1	O20217
4.1	1548	1	V55840
4.1	1550	1	O31550
4.1	1550	1	V41719
4.0	3198	1	X02974
4.0	3142	1	X27314
4.0	2638	1	V20468
4.0	1065	1	O25433
4.0	4100	1	O74052
4.0	4523	1	X00462
4.0	24379	1	T93095
4.0	24379	1	V25925
4.0	28604	1	T37329
4.0	28604	1	T92474

Accession	Protein	Length	Score	E-value	Database
C 35	Chromosomal fragment	33.4	4.0	28804	1 V81474
C 36	Sphingomonas 588 s	33.4	4.0	28804	1 V95812
C 37	DNA encoding human	3.9	740	1	027236
C 38	human hypochromic	3.9	1417	1	V27893
C 39	human alpha-1A adri	33	1776	1	Q86111
C 40	Human alpha-1A adri	33	1776	1	T11601
C 41	Human alpha-1 CDNF	33	1848	1	T97129
C 42	Alpha 1A adrenergic	33	2140	1	Q63180
C 43	Alpha 1A adrenergic	33	2140	1	T03130
C 44	Fucose dehydrogenase	33	2679	1	Q28895
C 45	Continuation (5 of	33	3.9	110000	1 V30459_4

ALIGNMENTS

RESULT	1
V31985	
ID	V31985 standard; cDNA; 842 BP

DR 26-SEP-1996: mouse cell/177.
 DE Mouse Down syndrome-cell adhesion molecule DS-CAM 5' cDNA.
 CC DS-CAM; Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KW holoprosencephaly; corpus callosum agenesis;
 KW schizencephaly; diagnosis; assay; mouse; ds; ss.
 OS Mus sp.
 PN W09817795-A1.
 PD 30-APR-1998.
 PF 23-OCT-1997; U19547.
 PR 25-OCT-1996; US-0293322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR.
 PT WPI: 98-271791/24.
 PR New isolated Down's Syndrome-cell adhesion molecule - used to
 PT develop products for detection, diagnosis and therapy of
 PT developmental and neurological abnormalities
 PS Claim 2; Page 80; 109pp; English.
 CC This cDNA sequence comprises the 5' region of a cDNA clone
 CC for murine Down syndrome-cell adhesion molecule (DS-CAM), a
 CC member of a novel subclass of the Ig superfamily with homology to
 CC neutral cell adhesion molecules. The middle region (see VJ1986)
 CC and 3' region (see VJ1987) of the clone are also provided. The
 CC murine DS-CAM clone was isolated from a C57 Black/6 mouse brain
 CC cDNA library using human DS-CAM cDNA (see VJ1981 and VJ1988) as
 CC probe. The invention also provides human DS-CAM proteins (see
 CC W42085-87), as well as expression vectors and host cells.
 CC transgenic animals, antisense oligonucleotides, and primers useful
 CC for amplification of DS-CAM nucleic acids. DS-CAM polypeptides are
 CC associated with developmental and neurological processes. They can
 CC be used in e.g. neural prosthetic devices used in entubulation
 CC methods of repairing (regenerating) damaged or severed peripheral
 CC nerves. The products can also be used in detection, diagnosis and
 CC therapy of developmental and neurological abnormalities such as
 CC Down syndrome, mental retardation, holoprosencephaly, agenesis of
 CC the corpus callosum, or schizencephaly. Antisense oligonucleotides
 CC are used for inhibiting translation of mRNA.
 SQ Sequence 842 BP; 182 A; 244 C; 238 G; 178 T;

Query Match	100.0%	Score 842;	DB 1;	length 842;
Best Local Similarity	100.0%	Pred. No. 2.1e-230;		
Matches 842;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

[illegible]

Db	121	TTCTCTGTGTCCAGAGCTTGGCAATGTTTTCAGTAAAGGCCCACTCAGCCTTACT	180
Oy	181	ttgtcatatgcatgcgttgcagagagtagtgtttgcaagcaccatcggggagcgtgtgtccct	240
Db	181	TTGTGCATGCATGCCTCAGAGAGTATGTTTTCAGAGCACATGCGGAGCGTGTGCTCCT	240
Oy	241	gcccgcgtgcagagcatctccctcgtgactctgaatgtaagtaacctaagagggcgagagaga	300
Db	241	GCCCCGTGCAGAGCATCCCTCTGTGACTCTCAGATGTATCTTAGACAGGGCGAGGAGAGA	300
Oy	301	ttctagatgtcccccggagatccgcacgtccatcccaatgycactctcccaattttccct	360
Db	301	TTCTACGATGTCCCGGGATTCGGCAGCTCCATCCCAATGGCACTTCCAAATTTTCCCT	360
Oy	361	ttccctctcaagcttcaagcacttaataatataactactatctgtcaacagctgaaa	420
Db	361	TTCTCCTTCAAGCTTCAGCACCTTAACTCATATATACTTACTATTTGCACGCGTAAA	420
Oy	421	acccttaaggaaaatatagaagtcagaaigtctccacatcaaeagctgttttaacggagacct	480
Db	421	ACCTTCAAGGAAATATTAAGATCAGATGTCCATCATCAAGGCTGTTTACGGGAGCCCT	480
Oy	481	atacagtcctgtgtgagagaccagaaaacca tgaagggcaatgtcgcggtgttcaagtga	540
Db	481	ATACAGTCCGTGTGAGGACACAGAAACCATGAGAGCATGTCCGCGTGTTCAAGTGCA	540
Oy	541	ttatccctcctcgtgtgagggcgtacgtctctgtctatgtgggaaagacaggttt	600
Db	541	TTATCCCTCTCGTGGTGGAGCGTACCTCTGTGTGCTATGAGGAAAGACACGGTTT	600
Oy	601	caactgtctcagagctcagatattctcacaatccacaggagagccttgaatataaagatg	660
Db	601	CACTTGTCTCGAGATCTAAGATTTCATCAATCCACGGGAGCCTGTATATTAAAGATG	660
Oy	661	ttcagaaacgaagatgaggcgtgtacaactaccgctgtgcatcg_ycggcagaagtttcgggggg	720
Db	661	TTTCAGAAACGAAGATGGCGCTGTACAACCTACCGCTGCATCGCGGCGACAGATTCGGCGGGG	720
Oy	721	agacagacacagagaactgtgcgcgagactgttgcgtgtcagaacacgacaaatcagccatc	780
Db	721	AGACGAGACAGAGCACTGCCGCGAGACTGTTCGTGTCAAGAACGCAACTCAGGCCATTC	780
Oy	781	catcctggaaggtttgaccacgcgcacaaaccatgagccgggacagcgtgtgagctgtgccttgc	840
Db	781	CATCCTGGAAGGTTTGACACCCACCAACCATGCGCGGACCGCTGGAGGCTGCTTTC	840
Oy	841	ca 842	
Db	841	CA 842	
RESULT 2			
v31988			
ID	v31988	standard: cDNA: 6413	BP.
AC	v31988:		
DT	28-SEP-1998	(first entry)	
DE	Human Down syndrome-cell adhesion molecule DS-CAN2 cDNA.		
KM	DS-CAN2: Down syndrome-cell adhesion molecule; neural cell;		
KW	signal transduction; trisomy 21; mental retardation;		
KM	holoprosencephaly; corpus callosum agenesis;		
OS	schizencephaly; diagnosis: assay; human; ds; ss.		
EH	Homo sapiens.		
FT	Key	Location/Qualifiers	
FT	CDS	453..5168	
FT		/*tag= a	
PN	MO9817795-A1.		
PD	30-APR-1998.		
PF	23-OCT-1997: U9547		
PA	25-OCT-1996: U9-029322.		
PA	(CEDA-) CEDARS SINAI MEDICAL CENT.		
P1	Korenberg JR:		
DR	WPI: 98-271791/24.		

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:44 ; Search time 4987.54 Seconds
(without alignments)
637.410 Million cell updates/sec

Title: US-08-956-991-7

Perfect score: 842

Sequence:

1 ccggcgccggcgccggcgag.....cgctggagctgcttgcga 842

Scoring table:

IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1.*

2: em_est2.*

3: em_est3.*

4: em_est4.*

5: em_est5.*

6: em_est6.*

7: em_est7.*

8: em_est8.*

9: em_est9.*

10: em_est10.*

11: em_est11.*

12: em_est12.*

13: em_est13.*

14: em_est14.*

15: em_est15.*

16: em_est16.*

17: em_est17.*

18: em_est18.*

19: em_est19.*

20: gb_est1.*

21: gb_est2.*

22: gb_est3.*

23: gb_est4.*

24: gb_est5.*

25: gb_est6.*

26: gb_est7.*

27: gb_est8.*

28: gb_est9.*

29: gb_est10.*

30: gb_est11.*

31: gb_est12.*

32: gb_est13.*

33: gb_est14.*

34: gb_est15.*

35: gb_est16.*

36: gb_est17.*

37: gb_est18.*

38: gb_est19.*

39: gb_est20.*

40: gb_est21.*

41: gb_est22.*

42: gb_est23.*

43: gb_est24.*

44: gb_est25.*

45: gb_est26.*
46: gb_est27.*
47: gb_est28.*
48: gb_est29.*
49: gb_est30.*
50: gb_est31.*
51: gb_est32.*
52: gb_est33.*
53: gb_est34.*
54: gb_est35.*
55: gb_est36.*
56: gb_est37.*
57: gb_est38.*
58: gb_est39.*
59: gb_est40.*
60: gb_est41.*
61: gb_est42.*
62: gb_est43.*
63: gb_est44.*
64: gb_est45.*
65: gb_est46.*
66: gb_est47.*
67: gb_est48.*
68: gb_est49.*
69: gb_est50.*
70: gb_est51.*
71: gb_est52.*
72: gb_est53.*
73: gb_est54.*
74: gb_est55.*
75: gb_est56.*
76: gb_est57.*
77: gb_est58.*
78: gb_est59.*
79: gb_est60.*
80: gb_est61.*
81: gb_est62.*
82: gb_est63.*
83: gb_est64.*
84: gb_est65.*
85: gb_est66.*
86: gb_est67.*
87: gb_est68.*
88: gb_est69.*
89: gb_est70.*
90: gb_est71.*
91: gb_est72.*
92: gb_est73.*
93: gb_est74.*
94: gb_est75.*
95: gb_est76.*
96: gb_est77.*
97: gb_est78.*
98: gb_est79.*
99: gb_est80.*
100: gb_est81.*
101: gb_est82.*
102: gb_est83.*
103: gb_est84.*
104: gb_est85.*
105: gb_est86.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	177.2	21.0	619	79	FR0002941	286724 F.rubripes
2	135	16.0	147	79	HSKC18B12	X88325 H.sapiens D

3	110.8	13.2	450	47	AI498283	tm85h11.x
4	65.4	7.8	522	79	AL027078	Fugu rubr
5	44	5.2	925	79	AL053013	Drosophil
6	43.4	5.2	1009	79	AL098882	Drosophil
7	42.8	5.1	910	79	AL065629	Drosophil
8	41.6	4.9	839	79	AL054280	Drosophil
9	41.4	4.9	641	45	AI357868	q13b02.x
10	41	4.9	788	79	AL100255	Drosophil
11	41	4.9	794	88	AL085798	q13b02.x
12	41	4.9	935	79	AL066051	Drosophil
13	40.2	4.8	1049	26	W42199	mc95c09.r1
14	39.8	4.7	587	42	AI107879	GH05755.5
15	39.8	4.7	925	79	AL053013	Drosophil
16	39.6	4.7	1064	80	AL107031	Drosophil
17	39.4	4.7	1101	79	AL071731	Drosophil
18	39	4.6	764	100	AQ327293	nbx50040P
19	38.6	4.6	964	79	AL052554	Drosophil
20	38.6	4.6	237	80	AL109149	Drosophil
21	38.6	4.6	472	79	AL103118	Drosophil
22	38.4	4.6	413	44	AI249175	q68g04.x
23	38.4	4.6	415	45	AI363984	q35b02.x
24	38.4	4.6	1201	80	AL103945	Drosophil
25	38.2	4.5	787	79	AL098749	Drosophil
26	38	4.5	469	45	AI356089	q164f03.x
27	38	4.5	1201	80	AL106555	Drosophil
28	37.8	4.5	776	79	AL099352	Drosophil
29	37.6	4.5	1101	79	AL071004	Drosophil
30	37.6	4.5	1101	80	AL108419	Drosophil
31	37.4	4.4	1201	80	AL106580	Drosophil
32	37.4	4.4	288	23	D60890	HUM136D06B
33	37.4	4.4	331	49	AI655949	tt41608.x
34	37.4	4.4	474	88	AQ866010	nbe0026L
35	37.4	4.4	1016	88	AQ894372	HS_4832_A
36	37.2	4.4	1101	79	AL108460	Drosophil
37	37.2	4.4	1101	80	AL060214	Drosophil
38	37.2	4.4	1101	80	AL108775	Drosophil
39	37	4.4	817	89	AQ896580	HS_5440_A
40	37	4.4	833	89	AQ896856	HS_5440_A
41	37	4.4	914	79	AL059740	Drosophil
42	37	4.4	918	79	AL053409	Drosophil
43	37	4.4	932	79	AL066742	Drosophil
44	37	4.4	947	79	AL077161	Drosophil
45	37	4.4	1101	79	AL072149	Drosophil
RESULT	1					
FR0002941						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
PRIMER:						
DESCR:						
One pass dye-terminator sequencing of osmid cloned genomic						
sequence.						
LOCATION/Qualifiers						
FEATURES						
source						
exon						
112	110.8	13.2	450	47	AI498283	tm85h11.x
113	65.4	7.8	522	79	AL027078	Fugu rubr
114	44	5.2	925	79	AL053013	Drosophil
115	43.4	5.2	1009	79	AL098882	Drosophil
116	42.8	5.1	910	79	AL065629	Drosophil
117	41.6	4.9	839	79	AL054280	Drosophil
118	41.4	4.9	641	45	AI357868	q13b02.x
119	41	4.9	788	79	AL100255	Drosophil
120	41	4.9	794	88	AL085798	q13b02.x
121	41	4.9	935	79	AL066051	Drosophil
122	40.2	4.8	1049	26	W42199	mc95c09.r1
123	39.8	4.7	587	42	AI107879	GH05755.5
124	39.8	4.7	925	79	AL053013	Drosophil
125	39.6	4.7	1064	80	AL107031	Drosophil
126	39.4	4.7	1101	79	AL071731	Drosophil
127	39	4.6	764	100	AQ327293	nbx50040P
128	38.6	4.6	964	79	AL052554	Drosophil
129	38.6	4.6	237	80	AL109149	Drosophil
130	38.6	4.6	472	79	AL103118	Drosophil
131	38.4	4.6	413	44	AI249175	q68g04.x
132	38.4	4.6	415	45	AI363984	q35b02.x
133	38.4	4.6	1201	80	AL103945	Drosophil
134	38.2	4.5	787	79	AL098749	Drosophil
135	38	4.5	469	45	AI356089	q164f03.x
136	38	4.5	1201	80	AL106555	Drosophil
137	37.8	4.5	776	79	AL099352	Drosophil
138	37.6	4.5	1101	79	AL071004	Drosophil
139	37.6	4.5	1101	80	AL108419	Drosophil
140	37.4	4.4	1201	80	AL106580	Drosophil
141	37.4	4.4	288	23	D60890	HUM136D06B
142	37.4	4.4	331	49	AI655949	tt41608.x
143	37.4	4.4	474	88	AQ866010	nbe0026L
144	37.4	4.4	1016	88	AQ894372	HS_4832_A
145	37.2	4.4	1101	79	AL108460	Drosophil
146	37.2	4.4	1101	80	AL060214	Drosophil
147	37.2	4.4	1101	80	AL108775	Drosophil
148	37	4.4	817	89	AQ896580	HS_5440_A
149	37	4.4	833	89	AQ896856	HS_5440_A
150	37	4.4	914	79	AL059740	Drosophil
151	37	4.4	918	79	AL053409	Drosophil
152	37	4.4	932	79	AL066742	Drosophil
153	37	4.4	947	79	AL077161	Drosophil
154	37	4.4	1101	79	AL072149	Drosophil
RESULT	1					
FR0002941						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
PRIMER:						
DESCR:						
One pass dye-terminator sequencing of osmid cloned genomic						
sequence.						
LOCATION/Qualifiers						
FEATURES						
source						
exon						
112	110.8	13.2	450	47	AI498283	tm85h11.x
113	65.4	7.8	522	79	AL027078	Fugu rubr
114	44	5.2	925	79	AL053013	Drosophil
115	43.4	5.2	1009	79	AL098882	Drosophil
116	42.8	5.1	910	79	AL065629	Drosophil
117	41.6	4.9	839	79	AL054280	Drosophil
118	41.4	4.9	641	45	AI357868	q13b02.x
119	41	4.9	788	79	AL100255	Drosophil
120	41	4.9	794	88	AL085798	q13b02.x
121	41	4.9	935	79	AL066051	Drosophil
122	40.2	4.8	1049	26	W42199	mc95c09.r1
123	39.8	4.7	587	42	AI107879	GH05755.5
124	39.8	4.7	925	79	AL053013	Drosophil
125	39.6	4.7	1064	80	AL107031	Drosophil
126	39.4	4.7	1101	79	AL071731	Drosophil
127	39	4.6	764	100	AQ327293	nbx50040P
128	38.6	4.6	964	79	AL052554	Drosophil
129	38.6	4.6	237	80	AL109149	Drosophil
130	38.6	4.6	472	79	AL103118	Drosophil
131	38.4	4.6	413	44	AI249175	q68g04.x
132	38.4	4.6	415	45	AI363984	q35b02.x
133	38.4	4.6	1201	80	AL103945	Drosophil
134	38.2	4.5	787	79	AL098749	Drosophil
135	38	4.5	469	45	AI356089	q164f03.x
136	38	4.5	1201	80	AL106555	Drosophil
137	37.8	4.5	776	79	AL099352	Drosophil
138	37.6	4.5	1101	79	AL071004	Drosophil
139	37.6	4.5	1101	80	AL108419	Drosophil
140	37.4	4.4	1201	80	AL106580	Drosophil
141	37.4	4.4	288	23	D60890	HUM136D06B
142	37.4	4.4	331	49	AI655949	tt41608.x
143	37.4	4.4	474	88	AQ866010	nbe0026L
144	37.4	4.4	1016	88	AQ894372	HS_4832_A
145	37.2	4.4	1101	79	AL108460	Drosophil
146	37.2	4.4	1101	80	AL060214	Drosophil
147	37.2	4.4	1101	80	AL108775	Drosophil
148	37	4.4	817	89	AQ896580	HS_5440_A
149	37	4.4	833	89	AQ896856	HS_5440_A
150	37	4.4	914	79	AL059740	Drosophil
151	37	4.4	918	79	AL053409	Drosophil
152	37	4.4	932	79	AL066742	Drosophil
153	37	4.4	947	79	AL077161	Drosophil
154	37	4.4	1101	79	AL072149	Drosophil
RESULT	1					
FR0002941						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
PRIMER:						
DESCR:						
One pass dye-terminator sequencing of osmid cloned genomic						
sequence.						
LOCATION/Qualifiers						
FEATURES						
source						
exon						
112	110.8	13.2	450	47	AI498283	tm85h11.x
113	65.4	7.8	522	79	AL027078	Fugu rubr
114	44	5.2	925	79	AL053013	Drosophil
115	43.4	5.2	1009	79	AL098882	Drosophil
116	42.8	5.1	910	79	AL065629	Drosophil
117	41.6	4.9	839	79	AL054280	Drosophil
118	41.4	4.9	641	45	AI357868	q13b02.x
119	41	4.9	788	79	AL100255	Drosophil
120	41	4.9	794	88	AL085798	q13b02.x
121	41	4.9	935	79	AL066051	Drosophil
122	40.2	4.8	1049	26	W42199	mc95c09.r1
123	39.8	4.7	587	42	AI107879	GH05755.5
124	39.8	4.7	925	79	AL053013	Drosophil
125	39.6	4.7	1064	80	AL107031	Drosophil
126	39.4	4.7	1101	79	AL071731	Drosophil
127	39	4.6	764	100	AQ327293	nbx50040P
128	38.6	4.6	964	79	AL052554	Drosophil
129	38.6	4.6	237	80	AL109149	Drosophil
130	38.6	4.6	472	79	AL103118	Drosophil
131	38.4	4.6	413	44	AI249175	q68g04.x
132	38.4	4.6	415	45	AI363984	q35b02.x
133	38.4	4.6	1201	80	AL103945	Drosophil
134	38.2	4.5	787	79	AL098749	Drosophil
135	38	4.5	469	45	AI356089	q164f03.x
136	38	4.5	1201			

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:47:15 : Search time 7810.03 Seconds
(without alignments)
-349.125 Million cell updates/sec

Title: US-08-956-991-8

Sequence: 898 1 tgcgcgcgcgttgcagacct.....gcgccacacgttcgtcta 898

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_om: *
5: gb_om: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pl2: *
10: gb_pl2: *
11: gb_pl3: *
12: gb_ro: *
13: gb_ro: *
14: gb_ro: *
15: gb_sy: *
16: gb_sy: *
17: gb_vl: *
18: em_fun: *
19: em_hum1: *
20: em_hum2: *
21: em_hum: *
22: em_om: *
23: em_om: *
24: em_om: *
25: em_om: *
26: em_om: *
27: em_om: *
28: em_om: *
29: em_om: *
30: em_om: *
31: em_om: *
32: gb_htg1: *
33: gb_htg2: *
34: gb_in1: *
35: gb_in2: *
36: gb_in2: *
37: em_ba1: *
38: em_ba2: *
39: em_hum3: *
40: gb_pr4: *
41: gb_htg3: *
42: gb_htg4: *
43: gb_htg5: *
44: gb_htg6: *

45: gb_htg7: *
46: em_htg1: *
47: em_htg2: *
48: em_htg3: *
49: em_hum5: *
50: gb_pl3: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	502	55.9	6110	11	AF023449	AF023449 Homo sapi
C 2	502	55.9	6413	11	AF023450	AF023450 Homo sapi
C 3	244.8	27.3	6025	10	AB032958	AB032958 Homo sapi
C 4	170	18.9	159424	11	AF064865	AF064865 Homo sapi
C 5	74	8.2	193789	32	AP000849	AP000849 Homo sapi
C 6	39	4.3	1648	34	PIHRP1	X01469 Plasmidium
C 7	38.8	4.3	1238	5	AR064010	AR064010 Sequence
C 8	38.8	4.3	1271	11	HS057099	U57099 Human APEG-
C 9	38.8	4.3	2793	5	AR064011	AR064011 Sequence
C 10	38	4.2	96870	41	AC009739	AC009739 Drosophila
C 11	37.8	4.2	10915	2	AE001900	AE001900 Deinococc
C 12	37.8	4.2	103445	32	AP000644	AP000644 Homo sapi
C 13	37.6	4.2	855	34	PFHRP1	M15317 P. lophurae
C 14	37.6	4.2	1472	3	HRSIGMR	L49414 Equus caball
C 15	37.2	4.1	92797	11	AF038458	AF038458 Homo sapi
C 16	37.2	4.1	148432	35	AC004642	AC004642 Drosophila
C 17	36.6	4.1	1408	7	UKBNIFD2	D26272 Unknown nit
C 18	36.4	4.1	1408	7	ABU08609	U08609 Atropa bell
C 19	36.4	4.1	140	7	NSU08616	U08616 Nolana spat
C 20	36.4	4.1	721	5	166494	166494 Sequence 14
C 21	36.2	4.0	1308	5	AR064008	AR064008 Sequence
C 22	36.2	4.0	1308	12	AR064007	AR064007 Sequence
C 23	36.2	4.0	1308	12	RN057097	U57097 Rattus norv
C 24	35.8	4.0	13971	4	CHKONNE	D16541 Chicken DNA
C 25	35.2	3.9	1345	7	STRCPBCL	L12658 Streptozia
C 26	35.2	3.9	1400	8	MGICPBCL	XS4345 M. macrophyl
C 27	35.2	3.9	1428	7	CYCPRBCL	L01900 Cytillia rac
C 28	35.2	3.9	1428	7	ENICPBCL	L12616 Enkianthus
C 29	35.2	3.9	1428	7	MAICPBCL	L12555 Magnolia hy
C 30	35.2	3.9	1428	7	MAICPBCL	L12656 Magnolia sa
C 31	35.2	3.9	1428	7	MAICPBCL	L12659 Michelia fi
C 32	35.2	3.9	1428	7	MAICPBCL	L12666 Talium ova
C 33	35.2	3.9	1428	7	MAICPBCL	Y00815 Human mRNA
C 34	35.2	3.9	7702	9	HSIARR	AL079156 Streptomy
C 35	35.2	3.9	44375	1	SCG9	U92813 Hirudo med
C 36	35	3.9	6164	35	HMU92813	AC012006 Homo sapi
C 37	35	3.9	38178	42	AC012006	AC004952 Homo sapi
C 38	35	3.9	137968	33	AC004952	AF123662 Zaluzians
C 39	34.8	3.9	1322	8	AF123662	AF161798 Orobanch
C 40	34.8	3.9	1389	8	AF161798	U08608 Anthocercis
C 41	34.8	3.9	1408	7	AVU08608	U08613 Lycium cest
C 42	34.8	3.9	1408	7	LCU08613	U08614 Mandragora
C 43	34.8	3.9	1408	7	MU08614	U08615 Nicandra ph
C 44	34.8	3.9	1408	7	NP008615	U25344 Drosophila
C 45	34.8	3.9	2001	34	DM025344	

ALIGNMENTS

RESULT 1
AF023449/c AF023449 6110 bp mRNA PRI 01-JUN-1998
LOCUS Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)
DEFINITION mRNA, partial cds.
ACCESSION AF023449
VERSION AF023449.1 GI:3169765

KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE	1 (bases 1 to 6110) Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N., Lyons, G.E. and Korenberg, J.R. DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a Down Syndrome Region and is Involved in the Development of the Nervous System
TITLE	Unpublished 2 (bases 1 to 6110) Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N., Lyons, G.E. and Korenberg, J.R. Direct Submission Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research Institute, 110 George Burns Road, Davis Building, Suite 2005, Los Angeles, CA 90048-1869, USA
JOURNAL	Location/Qualifiers
REFERENCE	1. 6110
AUTHORS	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21 (Lrisomy 21)" /map="21q22, between HMG14 and MX1" /tissue-type="brain" /dev_stage="14 weeks, fetal" /clone="CHD-42" /note="derived from alternately-spliced mRNA" <1. 6110 /gene="DSCAM" <1. 5691 /gene="DSCAM" /note="member of immunoglobulin superfamily; involved in nervous system development" /product="Down syndrome cell adhesion molecule" /codon_start=1 /protein_id="AAC17966.1" /db_xref="GI:3169766" /translation="VSEDLHSISIVFNASLQEVFASTGTGLVPCPAIGPPTLHM YLATGEILVDFGIRHVNGLTLOIFPPSPSFTLIHNTYCYCAENPSGKIRSDV HRAVLEPYTAVREDOKTRGNVAFKILISSSEAYITTVYSMEKDVYSLSGSRFL HISTGALYIKDVONEDGLNYRCHTRHRTGTROSNARCTVSPANSPBILDGFD HKKAMAGORVELPCALGHPEDYRMLKNNMPELSGRQKVTLLLENIPDSGS YVCSNRYGTAKVIGRLYVYKOPALISPRKYSVSGVSLSCVGTEDQELSMY RNEELINPKGNVITGILNHNIMDMHVSDDGAYCCFVRKXLSADQVYVLEDTG PKLISAFSEKVVSPAPVSLMCMVKTPTITMTLDDPILKSGSHRSQMTSEGN VVSYLNISSQVRDGVTRCTANNSAGVLYQARIYVRGPASIRPKNTTATAGDLY HCRVIGYPTYSIKWYKNSNLPFNHROVAFEWNGTLKLSDVQKEDESEYCNLYO POLSTQSVHIVKVPPIQPEFEPRESIGQRFVFCVVASDGLPTITWQDGRIP GSLGTIDNIDFTSLRISNLIMHNGNTCIARNEAAVEHQSOLIVPRFVYVOP RQDDGIYKAVILNCSAEGYPTIVWKFSGAGVQFOPIALNGIQLYSGLMSLIK HYVEBDSGYLVCKVNDVADVSXMYLTKIPAMITSPNTLATOGCKKMSGLAK GKPIYVMEKEDEIRIINPMARYLSTKVEGVITLQILPTVRDSEGFSCDAINS YGEDRGITQLVQEPDPEPEIKDYKARITITRMWGMGDSNPITGYDIECKNSDS WDSAORTKDVSPOLNSATIIDHPSSTYIRMTAKRIGKSESNELTTTABADAPG PQEVHLEPISSOSIRYTKAPKHLONGIRQYQREYSTGNGVSPENYALISVDSG DSEVYTLDNLFKFTQGLVQACNRAGTGSSOEITITLLEDPSPENYALIASP ESTISIMSTSKELNGIIOGFRIYMANDELBEIKNITTOQSLELDGIEKTN YSIOVLAFTRACGVRSEQIFTRKEDVPPGPAVAAAASAMVSMPLKNGI IRKTYFCHSPPIPVISFASPDSPSVRTPNLSNRQYSVWVAATASGRGSSII TRIEPLAARAILFTSGVTTPMKADVLCKKAVGDSPRVAKMKSNGSPNLTITD RSISFNGSEFIIRTKAEDSGYSCIANNNWSDEITLNLQVPPDOPLATYKSTS SSITLWLPDNGSSIRGYLIQYSEDNSSOMSPISPSERSYLENLCGWTAFST LTAONGVGPRISEIETAKLKEPOFSKDELFASINTVRNLINGMWDCPIYTS FTLEBPRTVMTAORTLSKSYLIYDQEAITWELQVCSNAGCAKEQKRNFTL NYDSSTIPLIKLVYVONEGLTVNKGILMVTTCIYVALLGLLVVRRRQOR LKRLDKASLAELMSKNTSRDLSKOOCOTLNMHIDIPRAOLLIERDTEMIDRS TVLLTDADFGAAOKSLTYHTVHTVSQATGCPLYDVSADSDSISVSQSOTDAR TARNRYASQWTLNRPHTISAHLLTWTWRLPTPRAASGVKESDSISVSQSOTDAR SSMSTESASGICIELARAYEHAKMEOLRHAFTITECFISDTSFQTLAGTNEYTD SLTSTPSESICRTFASPPKPDGGGVMNAVPAKAIQVTSYICLHTLEMTFC"
FEATURES	source
gene	
CDS	

BASE COUNT 1621 a 1658 c 1548 g 1283 t

ORIGIN

Query Match	55.9%	Score 502	DB 11	Length 6110
Best Local Similarity	82.3%	Pred. No. 8.2e-140		
Matches 749	Conservative	0	M.S.matches 135	Indels 26 Gaps 14
3	ccggcaggttgcaagccttacta--caagccatactgctgaa--ttatcaggtgtccca	59		
3169	CGGCCCGGTACAGGCGTGCACACAGGCGGACTGAGTAACATTATTCAGGTGTCCA	3110		
60	gggttgtaacctgcgtgt--ccgggtgtgtcaactgtagtgtgttgaacttggaagttac	118		
3109	GGGTGTAACCTCAGTGTGCCCCCTGTGTGCGACACTGATAATGTTGAATYGAAGTTAC	3050		
119	ccc-----gtgtctactccgtagcctatgtgagcgcgaatgatcccgctgtlatag	174		
3049	CCCCAGTGTGTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2990		
175	tggtcttggaagcctctccagatgaacctgatactctgagatgaggttggtccaaatga	234		
2989	GTTTCTTGGAAGCCT--TCCATGTGACCCCTGATCTGTGATGATGATGATGATGATG	2931		
235	actccctgaagtgga--cattacagagctgctccatccgctgtagtggctcgtctg	293		
2930	ACTTCTGAGGTGACATCAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2871		
294	ggtcactcttgccaatccggtctcttgctgctgatactgagatgctgtagtggaggaagg	353		
2870	GGCTCGCTCTGCGCAATCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2811		
354	tgaatatacagatggtgagcagatcagctgagtgagggaacattgtgtcttgagca	413		
2810	TGATATCAATGATGTGCGCGAGTTCAGCTGAGGGGAAACATCTTGGTCTCTGAGCA	2751		
414	gaatcccaagcttgatgattatttgatcattcaacacggtatagcctgtgtagtggctgt	473		
2750	GAAATCCAGAGGCTGAT TATTTTTGATT--CAATATGTAGCCGTATGAGGAGCTGT	2693		
474	tgccataaaccatcgtgtccactgagcgtgtagtggtagtggagcttgcaatccttgatc	533		
2692	TTCCATCAAAACCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2634		
534	tcaatcttggaagatcgtgaggtgtcttgacgtgtgagatgataatcctcaggtctcc	593		
2633	TCAATTTGGGAGGCTGTGGGCGCTTGTGACCTGTGACCTGATTAATTCAGGCTCCTCC	2574		
594	ccgataagattgatatgca tggcagagaagaagaacccggaatctctcctcaactgttgcaaa	653		
2573	CCATAGAGATTATATGATGCGAGGAAAGAAAGAAACCAATCTCTCACAGTTGGCAAA	2514		
654	atctgcaagctagatatacactctctctccacactcctctgtgtgatatata--gtacggcca	711		
2513	ACTGTCAGAGTGAATACCTCTTCTCCACCTCTGTTGAGGACACAAAGATTAACGGGCC	2454		
712	cttttagagttatgatactgtctctctct--ccagcggaatgatgaggtc--ctccca	768		
2453	ATCTCAGGTTATATATGATGCTGCTCTCTCTCTCCACGAGGATATAATGCGCTTCCACCA	2394		
769	tggtctgtgcaagctcat--tccctctcttgaccctgtagtggcaggtgtgtgtgataagt	827		
2393	TGGCGCGTACACTCATCTCTTTTTCGCCCTCGGCGGAGGCTGATTTTGGATAG	2334		
828	tatatcatgcccga-----attcccggtgagctcagctgagctgcaagctctgcgc	882		
2333	GATGTTATATGCGAGGAATTTAAACCGTGAAGTACATGAGACTTGTCTGA--CGTCTGGC	2276		
883	cccaatcgtt 892			
2275	CCCATCGTT 2266			

RESULT 2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:31:41 : Search time 383.18 Seconds
(without alignments)
586.337 Million cell updates/sec

Title: US-08-956-991-8
Perfect score: 898
Sequence: 1 tgcgcgcgttgcaagcct.....gcccacacgttcgtcgtga 898

Scoring table: IDENTITY_NUC
Gapcp 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	898	100.0	898	1	V31986 Mouse Down syndrom
2	502	55.9	6413	1	V31988 Human Down syndrom
3	502	55.9	6604	1	V31981 Human Down syndrom
4	38.8	4.3	1225	1	T60757 Human aortic prefe
5	38.8	4.3	1238	1	V48217 Human aortic-prefe
6	38.8	4.3	2793	1	V48218 Human striated mus
7	37.8	4.2	1308	1	Q25532 Sequence of genom
8	36.2	4.0	1308	1	V48215 Rat aortic prefe
9	36.2	4.0	1308	1	V48215 Rat aortic prefe
10	33.8	3.8	1032	1	V50486 Streptomyces clavu
11	33.8	3.8	6254	1	V60887 Coding sequence of
12	33.8	3.8	7193	1	V50431 Streptomyces clavu
13	33.8	3.8	9436	1	V63499 Blood transmissibl
14	32.6	3.6	2481	1	V23291 Synthetic HIV-1 gp
15	32.2	3.6	3003	1	X52268 Protein PRO334 (NP
16	32.2	3.6	3997	1	V62739 Human neural cell
17	31.8	3.5	1755	1	Q74310 Protein disulphide
18	31.8	3.5	1755	1	Q87670 Humicola insolens
19	31.6	3.5	132	1	T27590 Novel growth facto
20	31.4	3.5	255	1	T24442 Human gene -ignatu
21	31.4	3.5	1141	1	Q05690 Serum somatotropin
22	31.4	3.5	1557	1	T70441 H. pylori AlpB gen
23	31.4	3.5	1557	1	V07248 DNA encoding Helic
24	31.4	3.5	2547	1	Q05691 Serum somatotropin
25	31.4	3.5	11356	1	T13586 Synppl20nm. Synthe
26	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
27	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
28	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
29	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
30	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
31	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
32	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
33	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe
34	31.4	3.5	1632	1	T73951 HIV-1 gp120 synthe

C 35	30.8	3.4	2870	1	V71912 S. cerevisiae KIP2
C 36	30.8	3.4	4057	1	T36899 Mouse neuron rest
C 37	30.8	3.4	6567	1	Q03324 Elmeria tenella ge
C 38	30.6	3.4	2001	1	T94200 Thermotoga neapol
C 39	30.6	3.4	4942	1	X00461 Human type VI aden
C 40	30.2	3.4	4546	1	X23520 Human kidney amino
C 41	30.2	3.3	1563	1	V03867 Mouse hepatoma der
C 42	29.8	3.3	1335	1	X02886 Human mACHR-6 CDNA
C 43	29.8	3.3	2614	1	V48219 Human striated mus
C 44	29.8	3.3	2689	1	X02885 Human mACHR-6 CDNA
C 45	29.8	3.3	9578	1	V52260 Streptococcus pneu

ALIGNMENTS

RESULT 1	
ID V31986	standard; cDNA: 898 BP.
AC V31986:	
AT 28-SEP-1998	(first entry)
DE Mouse Down syndrome-cell adhesion molecule DS-CAM mid cDNA.	
KM DS-CAM: Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KM holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; mouse; ds; ss.	
OS Mus sp.	
PN M0981795-A1.	
PD 30-APR-1998.	
PF 23-OCT-1997; U19547.	
PR 25-OCT-1996; US-029322.	
PA (CEDA-) CEDARS SINAI MEDICAL CENT.	
PI korenberg JR:	
DI WPI: 98-271791/24	
PT New isolated Down's Syndrome-cell adhesion molecule - used to	
PT develop products for detection, diagnosis and therapy of	
PT developmental and neurological abnormalities	
PS Claim 2: Page 81; 109p; English.	
CC This cDNA sequence comprises the middle region of a cDNA clone	
CC for murine Down syndrome-cell adhesion molecule (DS-CAM), a	
CC member of a novel subclass of the Ig superfamily with homology to	
CC neural cell adhesion molecules. The 5' region (see V31985)	
CC and 3' region (see V31987) of the clone are also provided. The	
CC murine DS-CAM clone was isolated from a C57 Black/6 mouse brain	
CC cDNA library using human DS-CAM cDNA (see V31981 and V31988) as	
CC probe. The invention also provides human DS-CAM proteins (see	
CC M42086-87), as well as expression vectors and host cells,	
CC transgenic animals, antisense oligonucleotides, and primers useful	
CC for amplification of DS-CAM nucleic acids. DS-CAM polypeptides are	
CC associated with developmental and neurological processes. They can	
CC be used in e.g. neural prosthetic devices used in entubulation	
CC methods of repairing (regenerating) damaged or severed peripheral	
CC nerves. The products can also be used in detection, diagnosis and	
CC therapy of developmental and neurological abnormalities such as	
CC Down syndrome, mental retardation, holoprosencephaly, agenesis of	
CC the corpus callosum, or schizencephaly. Antisense oligonucleotides	
CC are used for inhibiting translation of mRNA.	
Sequence 898 BP; 174 A; 222 C; 245 G; 257 T;	

Query Match 100.0%; Score 898; DB 1; Length 898;
Best Local Similarity 100.0%; Pred. No. 1e-271;
Matches 898; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1	tgcgcgcgttgcaagccttactacagcgaacgcgtgaattacagttgtccag 60
DB 1	1 tgcgcgcgcgttgcaagccttactacagcgaacgcgtgaattacagttgtccag 60
QY 61	ggtgtacacattcgtcgtccggtgtgtcgaatcgtatgttgaacgtgaattccc 120
DB 61	ggtgtacacattcgtcgtccggtgtgtcgaatcgtatgttgaacgtgaattccc 120
QY 121	ctgtgtacacattcgtcgtccggtgtgtcgaatcgtatgttgaacgtgaattccc 180

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:50 ; Search time 4987.54 Seconds
(without alignments)
679.803 Million cell updates/sec

Title: US-08-956-991-8

Perfect score: 898
Sequence: 1 tgcgcgcgcgttcgaagcct.....gcgcacacatcgttcgtga 898

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 4538634 seqs, 1887831982 residue

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: gb_est1:*

21: gb_est2:*

22: gb_est3:*

23: gb_est4:*

24: gb_est5:*

25: gb_est6:*

26: gb_est7:*

27: gb_est8:*

28: gb_est9:*

29: gb_est10:*

30: gb_est11:*

31: gb_est12:*

32: gb_est13:*

33: gb_est14:*

34: gb_est15:*

35: gb_est16:*

36: gb_est17:*

37: gb_est18:*

38: gb_est19:*

39: gb_est20:*

40: gb_est21:*

41: gb_est22:*

42: gb_est23:*

43: gb_est24:*

44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: gb_est20:*

53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

59: gb_est33:*

60: gb_est34:*

61: gb_est35:*

62: gb_est36:*

63: gb_est37:*

64: gb_est38:*

65: em_est27:*

66: em_est28:*

67: em_est29:*

68: em_est30:*

69: gb_est39:*

70: gb_est40:*

71: gb_est41:*

72: gb_est42:*

73: gb_est43:*

74: gb_est44:*

75: em_est31:*

76: em_est32:*

77: em_est33:*

78: em_est34:*

79: gb_gss1:*

80: gb_gss2:*

81: gb_gss3:*

82: gb_gss4:*

83: em_gss1:*

84: em_gss2:*

85: em_gss3:*

86: em_gss4:*

87: gb_gss5:*

88: gb_gss6:*

89: gb_gss7:*

90: gb_gss8:*

91: gb_gss9:*

92: em_gss5:*

93: em_gss6:*

94: em_gss7:*

95: em_gss8:*

96: em_gss9:*

97: em_gss10:*

98: em_gss11:*

99: gb_gss10:*

100: gb_gss11:*

101: em_gss12:*

102: gb_gss12:*

103: gb_gss13:*

104: gb_gss14:*

105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	58	6.5	605	79	FR0021994	AL014865 F.rubripe
2	58	6.5	618	79	FR0022042	AL014913 F.rubripe

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:51:16 : Search time 7810.03 Seconds
(without alignments)
-844.819 Million cell updates/sec

US-08-956-991-9

Title: 2173
Perfect score: 1 accacattcacacaccacg.....aagaattgcacaatatatta 2173
Sequence:

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pl4:*
11: gb_pl5:*
12: gb_pl6:*
13: gb_pl7:*
14: gb_pl8:*
15: gb_pl9:*
16: gb_pl10:*
17: gb_pl11:*
18: gb_pl12:*
19: gb_pl13:*
20: gb_pl14:*
21: gb_pl15:*
22: gb_pl16:*
23: gb_pl17:*
24: gb_pl18:*
25: gb_pl19:*
26: gb_pl20:*
27: gb_pl21:*
28: gb_pl22:*
29: gb_pl23:*
30: gb_pl24:*
31: gb_pl25:*
32: gb_pl26:*
33: gb_pl27:*
34: gb_pl28:*
35: gb_pl29:*
36: gb_pl30:*
37: gb_pl31:*
38: gb_pl32:*
39: gb_pl33:*
40: gb_pl34:*
41: gb_pl35:*
42: gb_pl36:*
43: gb_pl37:*
44: gb_pl38:*
45: gb_pl39:*

45: gb_pl39:*
46: em_hlg1:*
47: em_hlg2:*
48: em_hlg3:*
49: em_hlg4:*
50: gb_pl39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1147.6	52.8	6110	11 AF023449	AF023449 Homo sapi
2	792.8	36.5	6413	11 AF023450	AF023450 Homo sapi
3	394.2	18.1	6025	10 AB032958	AB032958 Homo sapi
4	319.2	14.7	721	11 HUMZ91F03	HUMZ91F03 Homo sapi
5	236.8	10.9	1950	12 RNY15054	RNY15054 Rattus norv
6	230.6	10.6	145861	11 AF064862	AF064862 Homo sapi
7	230.6	10.6	159424	11 AF064865	AF064865 Homo sapi
8	154.2	7.1	430	13 G36681	G36681 SHGC-53823
9	109.6	5.0	36156	11 AC004789	AC004789 Homo sapi
10	109.6	5.0	40619	11 AC005222	AC005222 Homo sapi
11	109.6	5.0	172042	43 AC012676	AC012676 Homo sapi
12	81.6	3.8	96599	43 AC013845	AC013845 Drosophila
13	51.8	2.4	2658	8 CR078547	CR078547 Chlamydomon
14	47.6	2.2	63751	45 AC017489	AC017489 Drosophila
15	47.6	2.2	80446	35 AC004288	AC004288 Drosophila
16	45.2	2.1	196667	42 AC012068	AC012068 Homo sapi
17	44.6	2.1	1853	10 HUM56KDPAR	HUM56KDPAR Homo sapi
18	42.4	2.0	7218	5 I66494	I66494 Sequence 14
19	42.2	1.9	2250	4 AF069737	AF069737 Xenopus 1
20	41.4	1.9	2094C	43 AC014917	AC014917 Drosophila
21	41.2	1.9	2154	35 AF101480	AF101480 Trypanoso
22	41.2	1.9	32799	8 SPAC2227	SPAC2227 Arabidops
23	41.2	1.9	91916	8 AC005964	AC005964 Arabidops
24	41.2	1.9	110680	8 AC006259	AC006259 Arabidops
25	41.2	1.9	38666	7 SPAC24C9	SPAC24C9 S. pombe chr
26	40.8	1.9	35805	8 YSC19576	YSC19576 Saccharomyc
27	40.6	1.9	1886	12 AF143957	AF143957 Mus muscu
28	40.2	1.8	8665	11 HS086136	HS086136 Human telom
29	39.8	1.8	92586	33 AC009578	AC009578 Human telom
30	39.2	1.8	1428	5 E14330	E14330 cDNA encodi
31	39.2	1.8	1602	10 AB010098	AB010098 Homo sapi
32	39.2	1.8	3269	9 AB023142	AB023142 Homo sapi
33	39.2	1.8	46626	4 HSA243007	HSA243007 Homo sapi
34	38.6	1.8	1250	35 AF098070	AF098070 Drosophila
35	38.6	1.8	1756	35 AF152419	AF152419 Drosophila
36	38.6	1.8	2884	35 AF117606	AF117606 Drosophila
37	38.6	1.8	4586	10 HSM801203	HSM801203 Homo sapi
38	38.6	1.8	63433	34 AC004320	AC004320 Drosophila
39	38.6	1.8	11336	45 AC018305	AC018305 Drosophila
40	38.6	1.8	11222	12 MPURSD87	MPURSD87 Micromonosp
41	38.2	1.7	291288	42 AC008878	AC008878 Homo sapi
42	38.2	1.7	2552	34 DME243599	DME243599 Drosophila
43	37.8	1.7	38406	43 AC013837	AC013837 Drosophila
44	37.8	1.7	125013	32 DMBR11L10	DMBR11L10 Drosophila
45	37.8	1.7	125013	32 DMBR11L10	DMBR11L10 Drosophila

ALIGNMENTS

RESULT 1
AF023449 6110 bp mRNA
LOCUS Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)
DEFINITION
ACCESSION AF023449
VERSION AF023449.1 GI:3169765

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:32:03 ; Search time: 383.18 Seconds
(without alignments)
1418.830 Million cell updates/sec

Title: US-08-956-991-9

Perfect score: 2173
Sequence: 1 accaccattcacacacccag.....aagaatgcccaatatatta 2173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2165.6	99.7	2173	V31987	Mouse Down syndrom
2	1147.6	52.8	6604	V31981	Human Down syndrom
3	792.8	36.5	6413	V31988	Human Down syndrom
4	304.2	14.0	388	V31982	Human Down syndrom
5	40.2	1.8	7881	V27865	Human telomerase R
6	39.2	1.8	1428	V05861	Human neuronal str
7	38.6	1.8	8839	V25987	Human telomerase e
8	37.6	1.7	7042	V84785	Apoptosis inducer
9	37.6	1.7	7075	V84798	Apoptosis inducer
10	36.2	1.7	270	V87625	EST clone D1106. N
11	33.4	1.6	1171	Q13429	Calcium channel ga
12	33.4	1.5	8214	V25987	Rat telomerase enc
13	32.8	1.5	585	V19497	Cytochrome P450I1d
14	32.8	1.5	1494	O87730	Human auxillary cy
15	32.8	1.5	1494	O87731	Human auxillary cy
16	32.8	1.5	1494	O87732	Human auxillary cy
17	32.8	1.5	1494	O87732	Human auxillary cy
18	32.8	1.5	1494	T17388	Human derived cyto
19	32.8	1.5	1494	T28395	Human cytochrome P
20	32.8	1.5	1494	T28396	Human cytochrome P
21	32.8	1.5	1494	T28397	Human cytochrome P
22	32.8	1.5	1494	T28398	Human cytochrome P
23	32.8	1.5	1545	V19496	Cytochrome P450I1d
24	32.8	1.5	2152	O70732	TATA-binding prote
25	32.8	1.5	2152	T42218	Human TATA-binding
26	32.8	1.5	2152	T79594	TATA-binding prote
27	32.8	1.5	1396	O98498	Human 50 kDa dystr
28	32.6	1.5	1396	T97324	Human (50 kDa) cys
29	32.6	1.5	1436	V15734	Alpha-sarcoglycan
30	32.6	1.5	1436	V17873	Homo sapiens alpha
31	32.4	1.5	3670	O22501	Human MIR gene. Ma
32	32.2	1.5	2085	O81504	ORDE 45 kDa subuni
33	32.2	1.5	5496	X07327	Aspergillus oryzae
34	32	1.5	1506	Q47667	Sequence of the ra

35	32	1.5	1918	1	V54122	Mouse membrane pro
36	32	1.5	2060	1	T71579	Exon 5 of LETO rat
37	32	1.5	2187	1	V61810	Glucocorticoid ind
38	32	1.5	2217	1	X28434	EGF-like homologue
39	32	1.5	2217	1	X52215	Protein PRO230 CDN
40	31.8	1.5	543	1	X52215	Antigen tc-7a gene
41	31.6	1.5	1713	1	X20661	Poly nucleotide seq
42	31.6	1.5	2514	1	T30308	Mouse Sox-9 cDNA.
43	31.6	1.5	3256	1	V29063	BRCA1 modulator pr
44	31.6	1.5	3934	1	T30309	Human SOX-9 cDNA.
45	31.6	1.5	15894	1	V18264	Measles virus Edmo

ALIGNMENTS

RESULT 1	V31987	standard; cDNA; 2173 BP.
ID	V31987	
AC	V31987	
DT	28-SEP-1998	(first entry)
DE	Mouse Down syndrome-cell adhesion molecule DS-CAM 3' cDNA.	
KW	DS-CAM; Down syndrome-cell adhesion molecule; neural cell;	
KW	signal transduction; trisomy 21; mental retardation;	
KW	holoprosencephaly; corpus callosum agenesis;	
KW	schizencephaly; diagnosis; assay; mouse; ds; ss.	
OS	Mus sp.	
PN	W09817795-AL.	
PD	30-APR-1998.	
PF	25-OCT-1997; U19547.	
PR	25-OCT-1996; US-029322.	
PA	(CEDA-) CEDARS SINAI MEDICAL CENT.	
PI	Korenberg JR.	
DR	WPI: 98-271791/24.	
PT	New isolated Down's Syndrome-cell adhesion molecule - used to	
PT	develop products for detection, diagnosis and therapy of	
PT	developmental and neurological abnormalities	
PS	Claim 2; Page 81-83; 109pp; English.	
CC	This cDNA sequence comprises the 3' region of a cDNA clone	
CC	for murine Down syndrome-cell adhesion molecule (DS-CAM), a	
CC	member of a novel subclass of the Ig superfamily with homology to	
CC	neural cell adhesion molecules. The middle region (see V31986)	
CC	and 5' region (see V31985) of the clone are also provided. The	
CC	murine DS-CAM clone was isolated from a C57 Black/6 mouse brain	
CC	cDNA library using human DS-CAM cDNA (see V31981 and V31988) as	
CC	probe. The invention also provides human DS-CAM proteins (see	
CC	W42086-87), as well as expression vectors and host cells.	
CC	transgenic animals, antisense oligonucleotides, and primers useful	
CC	for amplification of DS-CAM nucleic acids. DS-CAM polypeptides are	
CC	associated with developmental and neurological processes. They can	
CC	be used in e.g. neural prosthetic devices used in entubulation	
CC	methods of repairing (regenerating) damaged or severed peripheral	
CC	nerves. The products can also be used in detection, diagnosis and	
CC	therapy of developmental and neurological abnormalities such as	
CC	Down syndrome, mental retardation, holoprosencephaly, agenesis of	
CC	the corpus callosum, or schizencephaly. Antisense oligonucleotides	
CC	are used for inhibiting translation of mRNA.	
SQ	Sequence 2173 BP; 585 A; 598 C; 554 G; 436 T;	
Query Match	99.7%; Score 2166.6; DB 1; Length 2173;	
Best Local Similarity	99.8%; Pred. No. 0;	
Matches 2169; Conservative	0; Mismatches 4; Indels 0; Gaps 0;	
QY	1 accaccattcacacccagacatgacgaggttcgcgcgaaccttcagtccttcgc 60	
DB	1 ACCACATTTCACACACCCAGACATGGCGGTTGCGCGCAACTTACGTTCTCGCCTTC 60	
QY	61 tgaaggtaaaagggctgctgctggtttatagacggcgacatgccatctctgcatacgt 120	
DB	61 TGAAGGTAAGGGCTGCTGCTGTTATAGACCGGCGACATGCCCATCTCTGCATACGCT 120	
QY	121 ggcacagtggttcctcatctgattccagcgccaagtaaaatctgtctgtgcccctg 180	

```
|||||
Db 121 GGGCAGTGGCTTTCATCTTGATTTCCAGGCCAAGTAAATCTGTCTGATGGCCCTG 180
Oy 181 cagtttaagccgtttcagctcccaagttcgaagttcccaagtcgaagcggttagatcataagga 240
Db 181 cagtttaagccgtttcagctcccaagttcgaagttcccaagtcgaagcggttagatcataagga 240
Oy 241 acvtgaagccagttcaacgcgcagcagcgggttgaaagcgcagagatagatcttctctgt 300
Db 241 ACTGGAAGCCAGTACATCGGAGCCAGGGGCTGGAAGCCGAGAGAGATGATCTTTCTGT 300
Oy 301 gttgcccgtgagcacaagttcgaagttgtgtgaagacattcgaagcagcagcgttcac 360
Db 301 GTGGCCTGTGAGCAGCTTCAGCTGTGTGAGAACATTCCTGAGCCAGCAGCGTTTCAT 360
Oy 361 accgcctggaacacctaaagtgtgtgaagttgtatgaagttcaccttaccgtcccaaatg 420
Db 361 ACCGCTTGGAACCTTAAGTGTGGACTTGTGATTAAGTTACCTTACTGCCCCAAATG 420
Oy 421 gagtaagtcctggcgcgcaaaagtaaatatagaaagccaatgccctgtggggaagaaacccc 480
Db 421 GAGTAGTCCCGGGCGGCAATAGTGAATCATAGAACCCAAACCCCTGGGGAAAGAACCCC 480
Oy 481 agttcccaaggaagcagagcttctgcgaacatcaatacaccgcagagtgaggtgagatc 540
Db 481 AGTTCTCCAGAGGAGCAGAGCTTTTCGCCACATCAATACCAACCGAGTGAAGCTGATC 540
Oy 541 tgatgagtcgaagtaagcagcgtgtgccatcaccttcaacttcaacttgaatataagacct 600
Db 541 TGATTGCTGGAATACAGCGGGCTGTCCATACCTCATTCACCTTGAATACAGACCT 600
Oy 601 ttggagcaacgggtctggaacacagctcagcggacctcccttccaaagtcctaaactctg 660
Db 601 TTGGAGCAACGGGTCTGGAGCACAGCTCAGCGGACTCCCTTCCAGTCTCAACTTCTG 660
Oy 661 tatgacctgcaagaaagcagcgtgtgataaactgacatgagtagtgcgaagcgcgcgc 720
Db 661 TATGACCTGCAAGAAAGCCAGTGTATGATGACGATGAGAGTGTGCACAGCGCCGCG 720
Oy 721 ttgtcgataaagcaagcaacttcgcacagctgaactgaatgagtagatcaactctctca 780
Db 721 TTGTCGGATTAACCAAGCCAACTTTCGCCACGCTGAACCTAGATGGAGTACATTCCTCA 780
Oy 781 ctcaatcaatcagttgttcacaaagcgaagaaagcgtgacaaaccaaagcgaagcctca 840
Db 781 CTCATTAAAGTCAAGTTGTCCACAAAGCGAAGAGGCTGACAAACGAAGGCTCAAGA 840
Oy 841 tctctgacacatctcctgcatcctgtgtcgggttctactgtcttctgtctgtcgtg 900
Db 841 TCCTCGTGAACATCTCTGCATCTGTGTGGGGTTCTACTGCTTTGTGCTTGTGCTG 900
Oy 901 ttgtcgaggaagagcagcagcgaagcgaagcgtgaagagcgtgagagaatgcaagagttag 960
Db 901 TTGTGGGAGAGACGCGGAGCAGAGAGGCTGAAGAGCTGAGAGATGCAAAAGATTGAG 960
Oy 961 ctgaatatgtccatgagcaaaacacacagcagcttcagataccttaagcaaaacagcagcaga 1020
Db 961 CTGAATATGCTATGAGCAAAACACACAGGACTTCAGATACCTTAAACCAACGACACAGA 1020
Oy 1021 ctttgaagatgacatctgatatcccaagcgtcagcttttgattgagaagagaagacaaa 1080
Db 1021 CTTTGAGAAATGCATTTGATTAACCCAGGCTCAGTTTGTGATTGAAGAGAGACAAA 1080
Oy 1081 ttgagacacatagatgacagcttcccaagctcctgtgtgacgagtgctgcagcgttcgagagcag 1140
Db 1081 TGGAGACATAGATGACCGCTCCACAGTCTGTGAGAGAGTCTGACTTGGGGAGGAG 1140
Oy 1141 ccaaacgaagatcactgacagtgatcaacacagtgcttcaacaaatcggttctcaagcca 1200
Db 1141 CCAAAAGAAATCACTGACAGTGTACACAGCGGTGATTAACCAATGGTGTCTCAGGCCA 1200
Oy 1201 ccgggcccctgtgagtgctcgcagtgctcgcaggaagaaatcccaacacagagaga 1260
|||||
```

```
Db 1201 CCGGGCCCTCTGTGATGTCTCCATGCTCGGCCAGGATGAATGCCACCAGCAGAGGA 1260
Oy 1261 atgcaaaagctgagccccacagcaggaacccggtacgcgcagccagttgagcgtctaagaagac 1320
Db 1261 ATGCAAAAGCTGAGACCCACAGCGAGAAACCGGTACGCCACCACATGTGAGCCTCAACAGAC 1320
Oy 1321 cccatcccaacatctctgcagacacccctcaacacagatagactgtctacaacagagctac 1380
Db 1321 CCCATCTTACCATCTCTGACACACACCTCACACAGATGAGACTGTCTACACAGGCTAC 1380
Oy 1381 agatcccgtagaagagagcagtagaacgctcagcccatcacaagacacagagagc 1440
Db 1381 AGATCCGTGACAGAGAGCGCAGTAGTACAGCGTCAAGCTTCAAGACACACAGCAGC 1440
Oy 1441 aagaagcagcagtgcttccacagaagaagtgcttcttcaacagaagaagcagcagccta 1500
Db 1441 AAGAAAGCAGCATGTCTCTCACAGAAAGTCTTCTTCACTACAGAGACTCCAGGCTTA 1500
Oy 1501 tgaacacgcgaagatgagaagcagctgagacatgccaagttcacatcacagagtgctt 1560
Db 1501 TGAACACGCGCAAGATGGAAGAGCAGCTGAGGATGCCAAGTTCAACATCACAGAGTCTT 1560
Oy 1561 cabatccgatacgtctcctcgagcagtgacgycgagacaatatgtagacgagacgtctg 1620
Db 1561 CATATCCGATAGCTCTCTCGACAGTGAAGGCGCAGACAAATGATGATACGAGACAGTCTG 1620
Oy 1621 actccagtaaccccttcagacatcgagatcgcagattcaatgcatctcccccacactcagg 1680
Db 1621 ACTCCAGTACCCCTTCAGATGGGATCGGATTCATGATTCATCTCCCCCAACTCTCAGG 1680
Oy 1681 atggaagcagtgtagaacatgagcgggttccaaagcccatcgccagcagcactcaaac 1740
Db 1681 ATGAGAGAGAGAGTGAATGCGGCTTCCAAAGGCCCATCGGCAAGGAGTCTATAC 1740
Oy 1741 ctgtctcatacctcagtagatctctgtttaaocgagcgcacaaagcaccagcagagactg 1800
Db 1741 CTGCTCATACCTACATGATGATTTCTTTAAACCGGGCGCACAGGACACAGAGACTG 1800
Oy 1801 agtttagaagcagtgctgtgaaaccccaagaagtcggaaccttgaaagcccccagcgtgc 1860
Db 1801 AGTTTAGACAAACGCTGTGAAACCCAGAAAGTGGACCTGAAACGCCCTGAAACGCCACGCTG 1860
Oy 1861 ttgagcccaaccccttagaagcgtctcctccactcttcacagcgcagagaagcagcagct 1920
Db 1861 TTGAGCCCAACCCCTTAGAGAGGCTTCTCTCCACTTCTTCCACGCGAAGAGAGAGCACT 1920
Oy 1921 cgtgcaacaagagctgtagccaccttcctcaycgaagaggtgcagagctgcagacgc 1980
Db 1921 CGTGGCAACAGGGGCTGTGGCCACCTTACTCAGCAGAGCGTGCAGAGCTGCAGACGC 1980
Oy 1981 agctaaatagcagctcccaagagtgatcgtgtgaactcccggtgcatatgaaagagaca 2040
Db 1981 AGCTTAAATGAGCGCTCCAGAGGTACTGTGAGTCTCCGGGCCATTGAAAGAAACA 2040
Oy 2041 tccctacgaacatctacaccttgatataacacatgagcactgaagcagcgtgtgaat 2100
Db 2041 TCCTACGCAAAATTTACCTTGTGATTAACATGACATGACACTGATGACACGCTTTTAAT 2100
Oy 2101 acaatlaagagccaatcaactcttcttatagaatccgatatataatuaaatt 2160
Db 2101 ACAATTAAGAGCCAAATCAACTCTTTTATGAAATTCGATATTTATTAATTAACAAAT 2160
Oy 2161 gccaaatata 2173
Db 2161 GCCAAATATATTA 2173
```

RESULT 2
V31981
ID V31981 standard: cDNA: 6604 BP.
AC V31981.
DT 28-SEP-1998 (first entry)
DE Human Down. vndrome-cell adhesion molecule DS-CAM1 cDNA.

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:24:56 ; Search time 216.06 seconds
(without alignments)
1204.103 Million cell updates/sec

Title: US-08-956-991-9
Perfect score: 2173
Sequence: 1 accaccatcacaccaccag.....aagaatgcccaatatatta 2173

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 211294 seqs, 59861574 residues
Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/prodata/1/ina/5C.COMB.seq: *
4: /cgn2_6/prodata/1/ina/5D.COMB.seq: *
5: /cgn2_6/prodata/1/ina/6.COMB.seq: *
6: /cgn2_6/prodata/1/ina/PCTUS9_CC4B.seq: *
7: /cgn2_6/prodata/1/ina/backfiles1.seq: *

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
C 1	42.4	2.0	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	40.2	1.8	7881	4 US-08-751-189-1	Sequence 1, Appl
C 3	40.2	1.8	7881	4 US-09-060-836-1	Sequence 1, Appl
C 4	33.8	1.6	1171	2 US-08-336-257A-1	Sequence 1, Appl
C 5	33.8	1.6	1171	7 5386025-1	Patent No. 5386025
C 6	32.8	1.5	2152	1 US-08-188-582-17	Sequence 17, Appl
C 7	32.8	1.5	2152	1 US-08-646-715-17	Sequence 17, Appl
C 8	32.6	1.5	1396	1 US-08-123-161A-11	Sequence 11, Appl
C 9	32.6	1.5	1396	1 US-08-463-278-11	Sequence 11, Appl
C 10	32.2	1.5	2085	3 US-08-283-917-8	Sequence 8, Appl
C 11	32.2	1.5	2085	3 US-08-961-716-8	Sequence 8, Appl
C 12	32.2	1.5	1506	1 US-07-937-609-13	Sequence 13, Appl
C 13	31.8	1.5	543	7 5482709-5	Sequence 13, Appl
C 14	31.8	1.5	543	7 5482709-5	Patent No. 5482709
C 15	31.8	1.5	543	7 5482709-5	Patent No. 5482709
C 16	31.6	1.5	3256	4 US-08-968-751-3	Sequence 3, Appl
C 17	31.4	1.4	3640	4 US-08-627-873-6	Sequence 6, Appl
C 18	31.4	1.4	26700	2 US-08-472-217-1	Sequence 1, Appl
C 19	31.2	1.4	1569	4 US-08-488-199-5	Sequence 23, Appl
C 20	31.2	1.4	2521	2 US-08-145-658D-23	Sequence 23, Appl
C 21	31.1	1.4	2521	4 US-08-368-803-16	Sequence 16, Appl
C 22	30.8	1.4	1050	2 US-08-578-096A-18	Sequence 18, Appl
C 23	30.8	1.4	1050	2 US-08-599-252-81	Sequence 81, Appl
C 24	30.8	1.4	1050	6 PCT-US96-06352-81	Sequence 54, Appl
C 25	30.6	1.4	1050	6 PCT-US96-06352-81	Sequence 81, Appl
C 26	30.6	1.4	2149	2 US-08-784-651-3	Sequence 3, Appl
C 27	30.4	1.4	1365	2 US-08-420-235B-32	Sequence 32, Appl

C 28	30.4	1.4	1365	6 PCT-US95-10194-32	Sequence 32, Appl
C 29	30.4	1.4	1501	4 US-08-145-658D-24	Sequence 24, Appl
C 30	30.4	1.4	1566	4 US-08-145-658D-13	Sequence 13, Appl
C 31	30.4	1.4	1701	1 US-07-863-169A-2	Sequence 2, Appl
C 32	30.4	1.4	1701	4 US-08-429-96A-2	Sequence 2, Appl
C 33	30.4	1.4	1701	6 PCT-US93-08062-2	Sequence 2, Appl
C 34	30.4	1.4	20710	6 US-08-420-235B-1	Sequence 1, Appl
C 35	30.4	1.4	20710	6 PCT-US95-10194-1	Sequence 1, Appl
C 36	30.4	1.4	35100	3 US-08-770-379-18	Sequence 18, Appl
C 37	30.2	1.4	1568	4 US-08-145-658D-20	Sequence 20, Appl
C 38	30	1.4	1563	4 US-08-743-637B-183	Sequence 183, Appl
C 39	30	1.4	2307	3 US-08-967-101-28	Sequence 28, Appl
C 40	30	1.4	2307	4 US-08-592-541-28	Sequence 28, Appl
C 41	30	1.4	2416	2 US-08-592-126-60	Sequence 60, Appl
C 42	30	1.4	2416	2 US-08-592-126-61	Sequence 61, Appl
C 43	30	1.4	2416	2 US-08-592-126-63	Sequence 63, Appl
C 44	30	1.4	3366	2 US-08-469-802B-1	Sequence 1, Appl
C 45	30	1.4	3366	3 US-08-267-803B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F.G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F15
US-08-232-463-14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:56 ; Search time 4987.54 Seconds
(without alignments)
1645.003 Million cell updates/sec

Title: US-08-956-991-9
Perfect score: 2173
Sequence: 1 accaccatcacacacccag.....aagaattgccaaatatatta 2173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	292.2	13.4	541	64	AW046213
2	239	11.0	310	21	F13426 HSC2XA021 n

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 21:06:07 ; Search time 7810.03 Seconds
(without alignments)
-2493.247 Million cell updates/sec

Title: US-08-956-991-10

Perfect score: 6413
Sequence: 1 ttagctgagcgccgagcagcgcg.....gaaatgtcccaaatatatt 6413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_ov:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*
42: gb_ov:*
43: gb_ov:*
44: gb_ov:*

45: gb_ov:*
46: gb_ov:*
47: gb_ov:*
48: gb_ov:*
49: gb_ov:*
50: gb_ov:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	ID	Description
1	6413	100.0	6413	11 AF023450	AF023450 Homo sapi
2	5718	89.2	6110	11 AF023449	AF023449 Homo sapi
3	1845.8	28.8	6025	10 AB032958	AB032958 Homo sapi
4	699	10.9	721	11 HUMV291F03	HUMV291F03 Homo sapi
5	495.4	7.7	109866	11 AF043945	AF043945 Homo sapi
6	469.4	7.3	145861	11 AF064862	AF064862 Homo sapi
7	358.8	5.6	1430	13 G36681	G36681 SHGC-53823
8	319.4	5.0	119678	11 AF064866	AF064866 Homo sapi
9	319.4	5.0	120007	11 AF064864	AF064864 Homo sapi
10	307.2	4.8	121019	11 AF042091	AF042091 Homo sapi
11	307.2	4.8	155407	11 AF042090	AF042090 Homo sapi
12	267.4	4.2	159424	11 AF064865	AF064865 Homo sapi
13	154.6	2.4	189607	32 AP000757	AP000757 Homo sapi
14	154.6	2.4	223724	32 AP000834	AP000834 Homo sapi
15	153.8	2.4	153789	32 AP000849	AP000849 Homo sapi
16	152.8	2.4	124395	42 AC013397	AC013397 Homo sapi
17	121.4	1.9	193789	32 AP000849	AP000849 Homo sapi
18	98	1.5	96870	41 AC009739	AC009739 Homo sapi
19	96.4	1.5	279137	41 AC008540	AC008540 Homo sapi
20	91.8	1.4	110000	45 AC017014_0	AC017014_0 Homo sapi
21	91.2	1.4	174707	45 AC017082	AC017082 Homo sapi
22	89.2	1.4	107475	45 AC017106	AC017106 Homo sapi
23	89	1.4	7606	34 DROLARW	DROLARW Homo sapi
24	89	1.4	90563	42 AC019667	AC019667 Homo sapi
25	89	1.4	90563	42 AC019667	AC019667 Homo sapi
26	88.4	1.4	110000	45 AC017014_1	AC017014_1 Homo sapi
27	87.2	1.4	87042	43 AC011885	AC011885 Homo sapi
28	86.6	1.4	20940	43 AC014917	AC014917 Homo sapi
29	86.6	1.4	110000	45 AC017014_1	AC017014_1 Homo sapi
30	86.6	1.4	141025	44 AC015693	AC015693 Homo sapi
31	86.4	1.3	181634	44 AC016253	AC016253 Homo sapi
32	86	1.3	96106	45 AC017015	AC017015 Homo sapi
33	83.4	1.3	162448	43 AC010827	AC010827 Homo sapi
34	83.2	1.3	147343	44 AC011871	AC011871 Homo sapi
35	83.2	1.3	173721	43 AC015866	AC015866 Homo sapi
36	83	1.3	107475	45 AC017106	AC017106 Homo sapi
37	82.8	1.3	840	8 CNSO1BNS	CNSO1BNS Homo sapi
38	81.4	1.3	56416	41 AC011152	AC011152 Homo sapi
39	80.4	1.3	182110	44 AC005302	AC005302 Mus muscu
40	80.4	1.3	194355	33 AL133216	AL133216 Homo sapi
41	80.2	1.3	139207	41 AC011429	AC011429 Homo sapi
42	80.2	1.3	159836	44 AC008334	AC008334 Drosophi
43	79.6	1.2	181634	44 AC016253	AC016253 Homo sapi
44	79.4	1.2	60639	43 AC015463	AC015463 Homo sapi
45	79.4	1.2	67220	45 AC017014_3	AC017014_3 Homo sapi

ALIGNMENTS

RESULT 1
AF023450 6413 bp mRNA
LOCUS AF023450
DEFINITION Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (DSCAM)
ACCESSION AF023450
VERSION AF023450.1 GI:3169767

KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE	1 (bases 1 to 6413) Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N., Lyons, G.E. and Kornberg, J.R. DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a Down Syndrome Region and is Involved in the Development of the Nervous System
TITLE	Unpubl:shed
JOURNAL	2 (bases 1 to 6413) Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N., Lyons, G.E. and Kornberg, J.R.
REFERENCE	Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research Institute, 110 George Burns Road, Davis Building, Suite 2005, Los Angeles, CA 90048-1869, USA
JOURNAL	Location/Qualifiers
FEATURES	1..6413 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21 (trisomy 21)" /map="21q22, between HMGL4 and MX1" /tissue="brain" /dev_stage="14 weeks, fetal" /clone="CHD2-52" /note="Derived from alternately-spliced mRNA"
source	1..6413 /gene="DSCAM" /gene="DSCAM" 453..5168 /note="member of immunoglobulin superfamily; involved in nervous system development" /codon_start=1 /product="Down syndrome cell adhesion molecule" /protein_id="GI:3169768" /translation="MILALSLFOSFANVPSEDIHSLYFVNASLOEVPFASSTGTLVPCPAGISPRVLTLMVLTAGEETIDPGRAHYHPTQIPIPPSPSRSLIHNTYYCTAENPSGIRSDQVHIAKALREPTVREDQKTRGNVAVKCIIPSSVAYITVYVSEEDTVSLVSGREFLITSGALYIKDVONEDGLYRCITRHRITGTEROSAKRLVTDANSAPSLIDGDFHRKAMAGORVELPCALGHEPEYRWLKDMPLELSGRQKVTGLIENIRPSDGSVYCEVSNRYGAKVIGRLYKOPLEKATISPRKXSSGVSYSLCSVTIGDEDELSTWRNGELINDGKNVITGINHNLIMDHVKDGGAYOCFPAKRLSAODIVOVVLEDTGPKIISATSEKVSAPBVSIMCNYKGPILITITWIDDDPLIKGGSIRISOMITSEGNVSYLNTSSQVRDGGVRCRTANNSAGVLYOARIVRGPASIRPKMNTAIGRDYIHCRTGYPIYSIKWYKNSLILPNHQVAFENNGTLKLSDNQK
gene	1..6413 /gene="DSCAM" /gene="DSCAM" 453..5168 /note="member of immunoglobulin superfamily; involved in nervous system development" /codon_start=1 /product="Down syndrome cell adhesion molecule" /protein_id="GI:3169768" /translation="MILALSLFOSFANVPSEDIHSLYFVNASLOEVPFASSTGTLVPCPAGISPRVLTLMVLTAGEETIDPGRAHYHPTQIPIPPSPSRSLIHNTYYCTAENPSGIRSDQVHIAKALREPTVREDQKTRGNVAVKCIIPSSVAYITVYVSEEDTVSLVSGREFLITSGALYIKDVONEDGLYRCITRHRITGTEROSAKRLVTDANSAPSLIDGDFHRKAMAGORVELPCALGHEPEYRWLKDMPLELSGRQKVTGLIENIRPSDGSVYCEVSNRYGAKVIGRLYKOPLEKATISPRKXSSGVSYSLCSVTIGDEDELSTWRNGELINDGKNVITGINHNLIMDHVKDGGAYOCFPAKRLSAODIVOVVLEDTGPKIISATSEKVSAPBVSIMCNYKGPILITITWIDDDPLIKGGSIRISOMITSEGNVSYLNTSSQVRDGGVRCRTANNSAGVLYOARIVRGPASIRPKMNTAIGRDYIHCRTGYPIYSIKWYKNSLILPNHQVAFENNGTLKLSDNQK
CDS	1..6413 /gene="DSCAM" /gene="DSCAM" 453..5168 /note="member of immunoglobulin superfamily; involved in nervous system development" /codon_start=1 /product="Down syndrome cell adhesion molecule" /protein_id="GI:3169768" /translation="MILALSLFOSFANVPSEDIHSLYFVNASLOEVPFASSTGTLVPCPAGISPRVLTLMVLTAGEETIDPGRAHYHPTQIPIPPSPSRSLIHNTYYCTAENPSGIRSDQVHIAKALREPTVREDQKTRGNVAVKCIIPSSVAYITVYVSEEDTVSLVSGREFLITSGALYIKDVONEDGLYRCITRHRITGTEROSAKRLVTDANSAPSLIDGDFHRKAMAGORVELPCALGHEPEYRWLKDMPLELSGRQKVTGLIENIRPSDGSVYCEVSNRYGAKVIGRLYKOPLEKATISPRKXSSGVSYSLCSVTIGDEDELSTWRNGELINDGKNVITGINHNLIMDHVKDGGAYOCFPAKRLSAODIVOVVLEDTGPKIISATSEKVSAPBVSIMCNYKGPILITITWIDDDPLIKGGSIRISOMITSEGNVSYLNTSSQVRDGGVRCRTANNSAGVLYOARIVRGPASIRPKMNTAIGRDYIHCRTGYPIYSIKWYKNSLILPNHQVAFENNGTLKLSDNQK
BASE COUNT	1633 a 1779 c 1709 g 1292 t
ORIGIN	100.0%; Score 6413; DB 11; Length 6413; Best Local Similarity 100.0%; Pred. No. 0;

Matches 6413: Conservative	0: Mismatches	0: Indels	0: Gaps	0:
Qy 1	1	1	1	1
Db 1	1	1	1	1
Qy 61	61	61	61	61
Db 61	61	61	61	61
Qy 121	121	121	121	121
Db 121	121	121	121	121
Qy 181	181	181	181	181
Db 181	181	181	181	181
Qy 241	241	241	241	241
Db 241	241	241	241	241
Qy 301	301	301	301	301
Db 301	301	301	301	301
Qy 361	361	361	361	361
Db 361	361	361	361	361
Qy 421	421	421	421	421
Db 421	421	421	421	421
Qy 481	481	481	481	481
Db 481	481	481	481	481
Qy 541	541	541	541	541
Db 541	541	541	541	541
Qy 601	601	601	601	601
Db 601	601	601	601	601
Qy 661	661	661	661	661
Db 661	661	661	661	661
Qy 721	721	721	721	721
Db 721	721	721	721	721
Qy 781	781	781	781	781
Db 781	781	781	781	781
Qy 841	841	841	841	841
Db 841	841	841	841	841
Qy 901	901	901	901	901
Db 901	901	901	901	901
Qy 961	961	961	961	961
Db 961	961	961	961	961
Qy 1021	1021	1021	1021	1021
Db 1021	1021	1021	1021	1021

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd
OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:33:06 : Search time 383.18 Seconds
(without alignments)
4187.280 Million cell updates/sec

Title:	US-08-956-991-10
Perfect score:	6413
Sequence:	1 tgaactgagccgagacag.....gaaattgccaaatatatt 6413

Scoring table: IDENTITY__NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170
```

```
Minimum DB seq length: 0
Maximum DB seq length: 1000000
```

post-processing: Minimum Match 08
Listing first 45 summaries

Database : N_Geneseq_36: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	DB	ID	Description
No.	Score					
1	6409.8	100.0	6413	1	V31988	Human Down syndrom
2	6207.2	96.8	6604	1	V31981	Human Down syndrom
3	786.4	12.3	2173	1	V31987	Mouse Down syndrom
4	670.2	10.5	842	1	V31985	Mouse Down syndrom
5	501	8.4	1493	1	V27207	CDNA clone eth001
6	542	7.8	898	1	V31986	Mouse Down syndrom
7	143.8	2.2	388	1	V31982	Human Down syndrom
8	79.6	1.2	114955	1	X53491	Human adenosine A1
9	77.4	1.2	270	1	V87625	EST clone DY106. N
10	76	1.2	114955	1	X53491	Human adenosine A1
11	73	1.1	13198	1	X02974	Human IL-1ra BAC C
12	66.4	1.0	6000	1	O86478	Human PTP-OB. Prot
13	66.4	1.0	6000	1	T85389	Human protein tyro
14	66.4	1.0	6000	1	X06095	Human protein tyro
15	62.2	1.0	801	1	V73801	KSHV LTR terminal
16	62.2	1.0	6225	1	X55273	Human enzyme-relat
17	62	1.0	4257	1	V10352	Infected cell prot
18	62	1.0	4257	1	V68520	The nucleotide seq
19	62	1.0	12001	1	O76213	HSV L/ST region. H
20	62	1.0	117213	1	V62176	HSV-2 strain SB5 C
21	61.6	1.0	2218	1	O47929	Paired basic amino
22	61.6	1.0	4403	1	O47927	Paired basic amino
23	60.4	0.9	4020	1	T91361	Orf virus genomic
24	60.4	0.9	8438	1	O73500	DNA encoding pseud
25	60.4	0.9	117213	1	V62176	HSV-2 strain SB5 C
26	60.2	0.9	1026	1	O66091	MUSDUXX. a sample
27	59.4	0.9	4257	1	V10362	Infected cell prot
28	59.4	0.9	4257	1	V68520	The nucleotide seq
29	59.4	0.9	12001	1	O76213	HSV L/ST region. H
30	58.8	0.9	44377	1	T78508	Platanoidae syntha
31	58.8	0.9	44377	1	T80414	Platanoidae syntha
32	58.4	0.9	2504	1	O26467	M. scrofulaceum str
33	58.4	0.9	3946	1	T79510	Mycobacterium tube
34	58.2	0.9	1266	1	X09011	Bm-3a polynucleot

C	35	58.2	0.9	2974	1	Q79630	Partial FRAXE regi
C	36	57.8	0.9	6413	1	V31988	Human Down syndrom
C	37	57.8	0.9	6604	1	V31981	Human Down syndrom
C	38	57.4	0.9	1743	1	T85537	hNET cDNA sequence
C	39	57.4	0.9	1743	1	V16306	cDNA encoding humam
C	40	57.4	0.9	1974	1	V16346	Nucleotide sequenc
C	41	57.4	0.9	6803	1	T85473	Genomic hNET sequ
C	42	57.4	0.9	6803	1	V16305	Genomic DNA encod
C	43	57.2	0.9	1548	1	V33277	Nucleotide sequenc
C	44	57.2	0.9	1581	1	V33278	Nucleotide sequenc
C	45	56.8	0.9	2538	1	X06987	Mouse neurogulin 1

ALIGNMENTS

Query Match	100.0%;	Score 6409.8;	DB 1;	Length 6413;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 6411; Conservative	0;	Mismatches	2;	Indels 0; Gaps 0

Oy	1	tgaactaagcccgagacacgagcaagaataagcttgcgccgcctgtctgtcgtatgacg	60
Db	1	tcacttagcccgagacacgagcaagaatagcttgcgccgcctgtctgtcgtatgagccg	60
Oy	61	aagatgaagacgtgagcaccaggaagccgtgtgcctgtcgtcatcttgagcgccagatgcag	120
Db	61	agagatagagcctgagcaccaggaagccgtgtgcctgtcgtcatcttgagcgccagatgcag	120
Oy	121	gaataagctacgcccgcgtctgtcgtctcttccacttgcagacctgcggggggccactg	180
Db	121	gattgacctacgcccgcctgtctgcgctcttcttccacttgcagacctgcggggggccactg	180
Oy	181	agctgaacgagcaacttgagcttcggccgagcagatgaggagcgcgagcccggaacagatg	240
Db	181	agcttgacgagcaacttgagcttcggccgagcagatgaggagcgcgagcccggaacagatg	240
Oy	301	ccgggcggcttaagcagcggcggcggcggcggcggcggcggcggcggcggcggcggcggc	360
Db	301	ccgggcggcctaaagcagcggcggcggcggcggcggcggcggcggcggcggcggcggcggc	360
Oy	361	caagttccggagcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg	420
Db	361	caagttccggagcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg	420
Oy	421	tcgctcggcggaagccggggagcgagcgagcgagcgatgtgatactgtctctctgtcttc	480
Db	421	tcgctcggcggaagccggggagcgagcgagcgagcgagcgatgtgatactgtctctctgtcttc	480
Oy	481	agagatctcgagaaagtcttcagatgaagacctacacttccagccctacttctgaatgcat	540
Db	481	agagatctcgagaaagtcttcagatgaagacctacacttccagccctacttctgaatgcat	540
Oy	541	ctctcgcaagagtagtcttgcagacacacacgagggagcttgtgtccctgcggcagcag	600
Db	541	ctctcgcaagagtagtcttgcagacacacacgagggagcttgtgtccctgcggcagcag	600
Oy	601	gualcctcctctgtgaacttccagatgtgtacttagccaagggcgagagagatctagaatgcc	660
Db	601	gualcctcctctgtgaacttccagatgtgtacttagccaagggcgagagagatctagaatgcc	660
Oy	661	ccggagatccgcacagctcccaaccccaacgagcactcccaatttccctcctccctctca	720
Db	661	ccggagatccgcacagctcccaaccccaacgagcactcccaatttccctcctccctctca	720
Oy	721	gcttcagtaaccttaatccatgataaactatlatatgcagcgtgaaaatcccttcagga	780
Db	721	gcttcagtaaccttaatccatgataaactatlatatgcagcgtgaaaatcccttcagga	780
Oy	781	aaatlaaagatcgagatgttccacatacaagaagcgtttlaagggagccatacagtcggt	840
Db	781	aaatlaaagatcgagatgttccacatacaagaagcgtttlaagggagccatacagtcggt	840
Oy	841	tgagagaccagaaaaacccatagaagcgaatgtttgcgttccaagtcaatlaaccctcct	900
Db	841	tgagagaccagaaaaacccatagaagcgaatgtttgcgttccaagtcaatlaaccctcct	900
Oy	901	cggtgagagcgatcatcactgtctgtctcattggagaaagacgttttccacttctcag	960
Db	901	cggtgagagcgatcatcactgtctgtctcattggagaaagacgttttccacttctcag	960
Oy	961	gacttaatttctcatcaatccacagggagccctgtataatlaaagatgtlaagaaatgaag	1020
Db	961	gacttaatttctcatcaatccacagggagccctgtataatlaaagatgtlaagaaatgaag	1020
Oy	1021	atgagatgtataactaccgcttgcatacgcgagcatctgataacaccggagaagagagcaga	1080
Db	1021	atgagatgtataactaccgcttgcatacgcgagcatctgataacaccggagaagagagcaga	1080

QY	1081	gcacagcgcagaactttttgtatcaagaccagcgaactcagcccatcatcttgatg	1140
Db	1081	GCACAGCGCGCAACTTTTGTATCAGACCCAGGAATCGACCCCATCTACTGGATG	1140
QY	1141	ggtctaacctctcaagcaccatggctctggcagcgctgtggaactcgtctgaagcgctcg	1200
Db	1141	GGTTCACCTCCCAAGCCTATGCTGGGCAACGCTGTGACCTGCTTGCAAAAGCGCTGG	1200
QY	1201	ggcaccctgaagccagatlaacgcctggccttgaagacacatgcccctggaaacttcaggga	1260
Db	1201	GGCACCCCTGAGCCAGATTACCCCTGGCTGGAAGGACAACTATCCCTGGAACTTTCAGGGA	1260
QY	1261	ggttcgaagaagaccgctgaaggggcctggcctatggaagaatctggccctctgactcaagca	1320
Db	1261	GGTTCGAAAGACCGGAGCGGGGCTGCTCATTTGAACATATTCGCCCTCGGACTCAAGCA	1320
QY	1321	gcatctgtctgaagctgccacagatacgaacctgcctaaagtgtatagccgcgcctgaag	1380
Db	1321	GCTATGTTTGAAGTGTCACAGATGAGCACTGCTAAAGTATATGCGCGGCTGTAC	1380
QY	1381	tgaaacggccactggaagccaccatcatcagctccaggaagaagttaaagaagctggtgaagcc	1440
Db	1381	TGAACAGCCACTGGAAGCCACCATCATGCTCCAGAAAGTTAAAGCAAGCGTGGGTAGCC	1440
QY	1441	aagtttcctctgcctgcagcgttgacagagaactgaagaccagaacctctcctgtatccga	1500
Db	1441	AAGTTCTCTGTCTGTGAGGTGACAGGAACGTGAGACACAGAACTCTCCTGTATCCGCA	1500
QY	1501	atggtgaaatcctcaaaccttgaaaaaattgtgaagatcaaggaatcaacgaagaaaac	1560
Db	1501	ATGGTGAAATCTCTCAACCCCTGAAAAAATGTAGGATATCAGAGGATCAACGACCAAAAC	1560
QY	1561	ttataatggatcaatctgtcaaaagctgaacgggggcacatcacgtgtcttgtgcgaag	1620
Db	1561	TTATATATGATCACAATGTGTAAAGTAGAGGGGCGCATACCAATGCTTGTGTGGCAAGG	1620
QY	1621	acaaagctgcctccgtcctcaagactatgtcagggtgtcccttgaagatggaactcccaaatla	1680
Db	1621	ACAAAGCTGTCGCGCTCAAGACTATGTCAAGTGTGCTCTTGAGATGTGAATCCCAAAATTA	1680
QY	1681	ttctgccttttagtgaagaaggtgtgaagctcagaggaagccggtttcccttatgtgcgaag	1740
Db	1681	TTCTGCCTTTTAGTGAAGAAGGTGTGATGTCAAGCAAGCGGTTTCCCTTATGTGTCAAGC	1740
QY	1741	tgaaggaagacactcttgcacacatcaactatggaaccttgaagcattgaaccgaattctcaag	1800
Db	1741	TGAAGGAACACTCTTGTCCATCACTACCTGCACCTGTGAGCATGACCCGATTTCTCAAGG	1800
QY	1801	gtggcagctcacgcacatccacatctgactcagttcugagaaggagtglttcagctactctga	1860
Db	1801	GTGGCAGCTCACGCAGATCAGCCAGATGATACGCTCGAGGGGAACGTGCTACGTACTCTGA	1860
QY	1861	acaatctcagatctccaggtctcggagacggggaggtctcaacgctgcgaactcgaacaactcgg	1920
Db	1861	ACATCTCCAGCTCCCAAGTCCGGAGCGGGAGGTCTACACCGCTGCACCTCCAAACAATCGG	1920
QY	1921	cggagagctgcctctgtaccagcgtcogaataaactgaagaggctgcgcgaagcattgcagcaa	1980
Db	1921	CGGAGAGTGTCTCTTACCAAGCTCTGATTAACCTAAGAGGGCTCTCAAGCATTTGCACCAA	1980
QY	1981	tgaaaaacatacagcagaatgacgaaggaaggaacataatactactcgtgtgatgtgct	2040
Db	1981	TGAAAACATCACGCATATGACAGAGGGAGCACATACATTACTGTGTGTGATTGCT	2040
QY	2041	atccgtatctacatlaaataatgylacagaagactctaacgtgtcttcctttcaacacagcgc	2100
Db	2041	ATCCGTATTACTCATTAATAATGTACAGAACTCTAATCTGCTTTCAACACACGCC	2100
QY	2101	aagtgacttttgagaacaatggaactcttaactttcaatitgtgaaaaagaagaagtggagcg	2160
Db	2101	AAGTGCAATTTTGAGAACATGGAATCTTTAACTTTCAATGTGTGCAAAAGGAATGGAGCG	2160
QY	2161	aaggggaatlaacgttgcaagcgtgtgtttlaaaccaaacattctcacacagccagcagatgctc	2220

 WIRE (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Apr 25 13:12:33 2000: MasPar time 43.18 Seconds
 861.849 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-956-991-11
 Description: (1-1571) from US08956991A.pep
 Perfect Score: 11189
 Sequence: 1 MWILASLFSOFANVFSDDL.....NSAGCAEKAQKAEARCKEFS 1571

Scoring table: PAM 150
 GAP 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: a-geneseq36
 1:geneseqp

Statistics: Mean 40.973; Variance 226.720; scale 0.181

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description	Pred. No.
1	11180	99.9	1571	1	M42087 Human Down syndrome-ce	0.00e+00
2	11119	99.4	1910	1	M42086 Human Down syndrome-ce	0.00e+00
3	2081	18.6	465	1	M55045 Neural adhesion molecu	1.81e-145
4	740	6.6	1447	1	R68553 Deleted in colorectal	4.41e-43
5	740	6.6	1728	1	R13144 Deleted in Colorectal	4.41e-43
6	631	5.6	1018	1	R87028 Human contactin	5.75e-35
7	629	5.6	1018	1	R63759 Human contactin (EMBL	8.08e-35
8	623	5.6	1182	1	M57900 Protein of clone CO722	2.28e-34
9	620	5.5	1257	1	M74152 Human LI cell adhesion	3.75e-34
10	606	5.4	1018	1	M06485 Rat contactin ligand f	4.08e-33
11	608	5.4	1028	1	M29667 Homo sapiens DL185_1 c	2.90e-33
12	590	5.3	761	1	R92255 Neural cell adhesion m	6.20e-32
13	588	5.3	1304	1	M59994 Human neural cell adhe	8.71e-32
14	569	5.1	1911	1	M94027 Human protein tyrosine	2.19e-30
15	569	5.1	1911	1	M27225 Human protein tyrosine	2.19e-30
16	569	5.1	1911	1	R71726 Human PTP-OB	2.19e-30
17	549	4.9	582	1	R92256 Neural cell adhesion m	6.48e-29
18	525	4.7	1070	1	M08747 Human colon carcinoma	3.74e-27
19	514	4.6	753	1	M83927 Human T85 protein	2.39e-26
20	486	4.3	1501	1	R72858 Rat receptor type-prot	2.64e-24
21	414	3.7	1496	1	M81030 Melanoma associated an	4.29e-19
22	394	3.5	400	1	R75203 Tyrosine phosphatase M	1.16e-17
23	353	3.2	1291	1	R75201 Tyrosine phosphatase M	9.62e-15

24	321	2.9	467	1	R84094	Nsk2 receptor with put	1.72e-12
25	319	2.9	467	1	M62575	Alternatively spliced	2.37e-12
26	319	2.9	475	1	R94982	Nsk2 extracellular dom	2.37e-12
27	322	2.9	478	1	R92718	Mouse muscle-localized	1.46e-12
28	322	2.9	860	1	R92716	Mouse muscle-localized	1.46e-12
29	319	2.9	863	1	M62569	Alternatively spliced	2.37e-12
30	319	2.9	867	1	M62583	Mouse receptor tyrosin	2.37e-12
31	322	2.9	868	1	R92717	Mouse muscle-localized	1.46e-12
32	320	2.9	868	1	M26507	Rat Dmk receptor	2.02e-12
33	320	2.9	868	1	M26610	Rat muscle-specific k1	2.02e-12
34	326	2.9	869	1	M26506	Human Dmk receptor	7.67e-13
35	326	2.9	869	1	M26611	Human muscle-specific	7.67e-13
36	319	2.9	871	1	R84087	Nsk2 receptor	2.37e-12
37	319	2.9	871	1	M62568	Mouse receptor tyrosin	2.37e-12
38	319	2.9	873	1	R84092	Nsk2 receptor with alt	2.37e-12
39	319	2.9	873	1	M62573	Alternatively spliced	2.37e-12
40	319	2.9	881	1	R84091	Nsk2 receptor with alt	2.37e-12
41	319	2.9	881	1	M62572	Mouse Nsk2 (alternativ	2.37e-12
42	320	2.9	1225	1	M52289	Homo sapiens cdo tumo	2.02e-12
43	330	2.9	1242	1	M52287	Rattus norvegicus cdo	4.02e-13
44	317	2.8	863	1	R84088	Nsk2 receptor with alt	3.27e-12
45	302	2.7	1091	1	M41641	Sequence used in detec	3.63e-11

ALIGNMENTS

RESULT 1
 ID M42087 standard: Protein: 1571 AA.
 AC M42087;
 DT 28-SEP-1998 (first entry)
 DE Human Down syndrome-cell adhesion molecule DS-CAM2.
 KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KW holoprosencephaly; corpus callosum agenesis;
 KW schizencephaly; diagnosis; assay; human.
 OS Homo sapiens.
 PN M09817795-A1.
 PD 30-APR-1998.
 PE 23-OCT-1997; U19547.
 PR 25-OCT-1996; US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR.
 DR WPI: 98-271791/24.
 DR N-PSDB: V31988.
 PT New isolated Down's Syndrome-cell adhesion molecule - used to
 develop products for detection, diagnosis and therapy of
 PT developmental and neurological abnormalities
 PS Claim 2; Page 90-95; 109pp; English.
 CC This polypeptide comprises Down syndrome-cell adhesion molecule
 CC DS-CAM2, an extracellular soluble protein belonging to a novel
 CC subclass of the Ig superfamily with highest homology to neural cell
 CC adhesion molecules. Its amino acid sequence was deduced from cDNA
 CC clones (see V31982) isolated from a trisomy 21 foetal brain library.
 CC It is a splice variant of membrane-bound DS-CAM1 (see M42086), and
 CC lacks the entire transmembrane domain of DS-CAM1. The invention
 CC provides human and murine DS-CAM nucleic acid sequences (see also
 CC V31981, V31985-87), expression vectors and host cells, transgenic
 CC animals, antibodies, antisense oligonucleotides, and primers
 CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated
 CC with developmental and neurological processes. They can be used in
 CC e.g. neural prosthetic devices used in entubulation methods of
 CC repairing (regenerating) damaged or severed peripheral nerves, and
 CC also in bioassays to identify agonists and antagonists. The products
 CC can also be used in detection, diagnosis and therapy of developmental
 CC and neurological abnormalities such as Down syndrome, mental
 CC retardation, holoprosencephaly, agenesis of the corpus callosum,
 CC or schizencephaly.
 CC Sequence 1571 AA.

Query Match 99.9%; Score 11180; DB 1; Length 1571;
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;
 Matches 1569; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

D	b	1	MMLALS	LFOSFANVSESDHSL	LYVNSL	DEVNPAST	GTGLVPCPAAG	IPPTVLR	RYL	60																																											
Q	y	1	MMIALSL	LFOSFANVSESDHSL	LYVNSL	DEVNPAST	GTGLVPCPAAG	IPPTVLR	RYL	60																																											
D	b	61	ATGEIYDV	GIKHHV	PNGLTQ	IFPP	SSFSFL	IDHNTY	YCTAENP	SGK	IRSDV	HKA	120																																								
Q	y	61	ATGEIYDV	GIKHHV	PNGLTQ	IFPP	SSFSFL	IDHNTY	YCTAENP	SGK	IRSDV	HKA	120																																								
D	b	121	VLEPPTV	RVEDOK	TMGN	VA	FKCI	IPSS	VEAYIT	VSWEK	DVSLV	SGRFL	IT	STGA	180																																						
Q	y	121	VLEPPTV	RVEDOK	TMGN	VA	FKCI	IPSS	VEAYIT	VSWEK	DVSLV	SGRFL	IT	STGA	180																																						
D	b	181	LYIKD	VONEGL	NYRC	IT	HHRT	GTEN	ROS	NSAR	L	VPSP	PANS	SAPS	I	LDG	FP	HRA	MA	Q	240																																
Q	y	181	LYIKD	VONEGL	NYRC	IT	HHRT	GTEN	ROS	NSAR	L	VPSP	PANS	SAPS	I	LDG	FP	HRA	MA	Q	240																																
D	b	241	RVELP	CKAL	HPED	YR	WMLK	DN	MPLE	SG	RF	OK	TV	TGL	LENI	IR	SD	SGS	Y	CEV	SN	R	G	300																													
Q	y	241	RVELP	CKAL	HPED	YR	WMLK	DN	MPLE	SG	RF	OK	TV	TGL	LENI	IR	SD	SGS	Y	CEV	SN	R	G	300																													
D	b	301	TAKYIG	LYKOP	LKAT	IS	PR	KV	SS	VS	VS	LS	CS	V	T	ED	EO	EL	SW	R	NG	E	L	IP	RG	N	V	360																									
Q	y	301	TAKYIG	LYKOP	LKAT	IS	PR	KV	SS	VS	VS	LS	CS	V	T	ED	EO	EL	SW	R	NG	E	L	IP	RG	N	V	360																									
D	b	361	RITG	IN	EN	LIM	DM	WKS	SD	GA	Y	Q	EV	R	K	D	S	A	O	D	Y	V	G	V	LE	D	E	T	P	K	I	SA	F	S	E	K	V	N	S	P	420												
Q	y	361	RITG	IN	EN	LIM	DM	WKS	SD	GA	Y	Q	EV	R	K	D	S	A	O	D	Y	V	G	V	LE	D	E	T	P	K	I	SA	F	S	E	K	V	N	S	P	420												
D	b	421	AEP	SL	MC	NK	G	P	L	P	I	T	Y	T	T	DD	P	L	G	G	S	H	R	I	S	O	M	T	S	G	N	L	S	T	S	S	O	V	R	G	480												
Q	y	421	AEP	SL	MC	NK	G	P	L	P	I	T	Y	T	T	DD	P	L	G	G	S	H	R	I	S	O	M	T	S	G	N	L	S	T	S	S	O	V	R	G	480												
D	b	481	VYRC	TAN	NS	AG	V	L	Y	O	A	R	I	W	R	G	P	A	S	I	R	E	K	N	T	A	I	A	G	R	D	T	I	H	C	R	V	I	G	P	Y	S	I	K	W	N	540						
Q	y	481	VYRC	TAN	NS	AG	V	L	Y	O	A	R	I	W	R	G	P	A	S	I	R	E	K	N	T	A	I	A	G	R	D	T	I	H	C	R	V	I	G	P	Y	S	I	K	W	N	540						
D	b	541	SNLL	P	NH	R	O	V	A	E	B	N	G	T	L	S	D	V	O	K	E	V	D	E	G	E	T	C	N	V	L	V	O	P	L	S	T	S	O	S	V	H	T	V	K	P	P	I	Q	600			
Q	y	541	SNLL	P	NH	R	O	V	A	E	B	N	G	T	L	S	D	V	O	K	E	V	D	E	G	E	T	C	N	V	L	V	O	P	L	S	T	S	O	S	V	H	T	V	K	P	P	I	Q	600			
D	b	601	P	E	E	P	R	S	I	O	G	R	V	E	P	C	V	V	V	S	G	D	L	P	I	T	M	O	K	G	R	I	P	S	L	E	V	T	I	D	N	I	D	F	T	S	L	I	S	N	L	S	660
Q	y	601	P	E	E	P	R	S	I	O	G	R	V	E	P	C	V	V	V	S	G	D	L	P	I	T	M	O	K	G	R	I	P	S	L	E	V	T	I														

QY	1081	TTTTLEDDVPSTPPENVOAIATSPESISISWSTLSKELNLILOGFEVYITWALMDGELGET	1140
Db	1141	KNITTTPOPSLELDGLEKRYNYVSIOVLAFETFRAGDGVRSDEQIFETRTKEDVPGCPAGYKAAAA	1200
QY	1141	KNITTTPOPSLELDGLEKRYNYVSIOVLAFETFRAGDGVRSDEQIFETRTKEDVPGCPAGYKAAAA	1200
Db	1201	SASWVFWSWLPPLKLGIIIRKTYVFCSSHYPPIVISEFEASPDSESYRIPLNLSRNRQYSVM	1260
QY	1201	SASWVFWSWLPPLKLGIIIRKTYVFCSSHYPPIVISEFEASPDSESYRIPLNLSRNRQYSVM	1260
Db	1261	VVANTSGRBNSSSEITVEPLAAPARILITFEGTYVTPMKRIVLPCKAVODPSAAVVM	1320
QY	1261	VVANTSGRBNSSSEITVEPLAAPARILITFEGTYVTPMKRIVLPCKAVODPSAAVVM	1320
Db	1321	KDSNGTPTSLVTIDGRSIFESNGSFIIRTVKAEDSGYVSCIANNNMGSDIILLNLQVYP	1380
QY	1321	KDSNGTPTSLVTIDGRSIFESNGSFIIRTVKAEDSGYVSCIANNNMGSDIILLNLQVYP	1380
Db	1381	DOPLRVJYSKTTSSSITLWSLPGDNGSSIRGYILOYSEDNSEQWGSFPIPSSENSRYLEN	1440
QY	1381	DOPLRVJYSKTTSSSITLWSLPGDNGSSIRGYILOYSEDNSEQWGSFPIPSSENSRYLEN	1440
Db	1441	LKCGTWKFFLTTONGVGPRISEIIEATTLCKEPOFSKEQCLFASINTTFRVRLTGMN	1500
QY	1441	LKCGTWKFFLTTONGVGPRISEIIEATTLCKEPOFSKEQCLFASINTTFRVRLTGMN	1500
Db	1501	DGCGPIISFTLEVRPGCTTWTWTAORTSLSKSYLIYDLLEATWYLELQWRVNSAGACAEQ	1560
QY	1501	DGCGPIISFTLEVRPGCTTWTWTAORTSLSKSYLIYDLLEATWYLELQWRVNSAGACAEQ	1560
Db	1561	AKEAARCKEFS 1571	
QY	1561	AKEAARCKEFS 1571	

	RESULT	2
ID	W42086 standard; Protein:	1910 AA.
AC	W42086;	
DJ	28-SEP-1998 (first entry)	
DE	Human Down syndrome-cell adhesion molecule DS-CAM1. DS-CAM1; Down syndrome-cell adhesion molecule; neural cell signal transduction; trisomy 21; mental retardation; holoprosencephaly; corpus callosum agenesis; schizencephaly; diagnosis; assay; human.	
KM	Homo sapiens.	
OS		
FH	Key	Location/Qualifiers
FT	Peptide	1..23
FT	/label= Stig_peptide	
FT	Protein	24..1910
FT	/label= Mal_protein	
FM	Domain	24..887
FT	/label= IG	
FT	/note= "Immunoglobulin type-C2 domain"	
FF	Domain	888..1594
FT	/label= Fbn	
FT	/note= "fibronectin type III domain"	
FT	Domain	1595..1616
FT	/label= Transmembrane	
FT	Domain	1617..1910
FT	/label= Cytoplasmic	
FT	Region	24..126
FT	/label= Ig1	
FT	Region	127..225
FT	/label= Ig2	
FT	Region	226..316
FT	/label= Ig3	
FT	Region	317..409
FT	/label= Ig4	
FT	Region	410..506
FT	/label= Ig5	
FT	Region	507..603
FT	/label= Ig6	
FT	Region	

MPSEARCH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

Mpsrch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Apr 25 13:19:32 2000: Maspar time 77.56 Seconds
955.483 Million cell updates/sec
Tabular output not generated.

Title: >US-08-956-991-11
(1-1571) from US08956991A.pap
Description: 11188
Perfect Score: 1 MWLALSLFQSFANVFSEDL.....NSACCAEKOAKBAARCKERS 1571
Sequence: 1 MWLALSLFQSFANVFSEDL.....NSACCAEKOAKBAARCKERS 1571

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172405 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 55.886; Variance 106.625; scale 0.524

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	11009	98.4	1896	2	T08851	Down syndrome cell ad	0.00e+00
2	809	7.2	1028	2	I58164	BIG-1 protein - rat	2.85e-135
3	786	7.0	1028	2	A53449	Plasmodium-associat	2.17e-130
4	740	6.6	1447	2	A54100	tumor suppressor prot	1.17e-120
5	744	6.6	7962	2	I38346	elastic titin - human	1.67e-121
6	720	6.4	1040	2	A49356	transient axonal glyco	1.93e-116
7	713	6.4	2232	2	T13924	sdh protein - fruit f	5.74e-115
8	703	6.3	1036	2	S22383	axonin 1 precursor -	7.28e-113
9	701	6.3	1040	2	A34695	axonal glycoprotein T	1.92e-112
10	707	6.3	1427	2	I51669	tumor suppressor - Af	1.05e-113
11	680	6.1	1259	2	A43425	Bravo/Nr-CAM cell adh	4.90e-108
12	677	6.1	1268	1	A39640	neural cell adhesion	2.08e-107
13	663	5.9	1010	2	J00094	Fl1 protein precursor	1.78e-104
14	663	5.9	1091	2	S01998	contactin precursor -	1.77e-104
15	657	5.7	1239	2	S36126	neural cell adhesion	4.77e-99
16	637	5.7	1260	1	S05479	neural cell adhesion	4.77e-99
17	640	5.7	1898	2	S46216	leukocyte antigen-rel	1.13e-99
18	631	5.6	1018	2	A54744	contactin 1 precursor	8.47e-98
19	629	5.6	1018	2	I37246	contactin - human	2.21e-97
20	631	5.6	1020	2	S05944	neural cell surface	8.47e-98
21	631	5.6	1021	2	A57112	contactin precursor -	8.47e-98
22	629	5.6	1651	2	T14160	transmembrane recepto	2.21e-97
23	622	5.6	1894	2	C54689	protein-tyrosine-phos	6.32e-96

24	613	5.5	1018	2	JC4211	neural adhesion prote	4.69e-94
25	620	5.5	1257	1	A41060	neural cell adhesion	1.65e-95
26	609	5.4	2029	1	TDPE1K	protein-tyrosine-phos	3.17e-93
27	598	5.3	1897	1	TDHULK	leukocyte antigen-rel	6.05e-91
28	587	5.2	1333	2	PN0568	connectin 3B - chick	1.14e-88
29	570	5.1	1907	2	S50893	protein-tyrosine-phos	3.70e-85
30	574	5.1	1912	2	A56178	protein-tyrosine-phos	5.54e-86
31	556	5.0	1091	1	J0CHNL	neural cell adhesion	2.82e-82
32	555	5.0	1239	1	A32579	neuroglian - fruit fl	4.54e-82
33	522	4.7	725	2	JB0099	neural cell adhesion	2.62e-75
34	527	4.7	1070	2	JC4593	protein-tyrosine kin	2.49e-76
35	524	4.7	1088	1	J0XJNL	neural cell adhesion	1.02e-75
36	523	4.7	1092	1	JN0635	neural cell adhesion	1.64e-75
37	521	4.7	1691	1	D54689	protein-tyrosine-phos	4.19e-75
38	518	4.6	725	2	JE0100	neural cell adhesion	1.72e-74
39	509	4.5	761	1	J0HUNG	neural cell adhesion	1.17e-72
40	498	4.5	853	1	J0B0NC	neural cell adhesion	2.02e-70
41	501	4.5	1265	1	A37967	neural cell adhesion	4.97e-71
42	505	4.5	1443	2	I50600	neogenin - chicken (f	7.63e-72
43	504	4.5	1863	2	S46217	protein-tyrosine-phos	1.22e-71
44	496	4.4	1115	1	J0MSNL	neural cell adhesion	5.14e-70
45	490	4.4	1344	2	T14316	rig-1 protein - mouse	8.48e-69

ALIGNMENTS

RESULT 1
ENTRY T08851 #type fragment
TITLE Down syndrome cell adhesion protein 1 - human (fragment)
ALTERNATE_NAMES Down syndrome cell adhesion molecule
ORGANISM #formal_name Homo sapiens #common_name man
11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

ACCESSIONS T08851
REFERENCE 216495
#authors Yamakawa, K.; Huo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.; Korenberg, J.R.

#submission submitted to the EMBL Data Library, September 1997
#description DSCAM: A novel member of the immunoglobulin superfamily maps in a down syndrome region and is involved in the development of the nervous system.

#accession T08851
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues 1-1896 ##label YAM
##cross-references EMBL:AF023449; NID:q3169765; PID:q3169766
##experimental_source brain: developmental stage: 14 weeks; fetal

GENETICS

#gene DSCAM
#map_position 21q22
#note derived from alternatively-spliced mRNA

FUNCTION
#description involved in nervous system development
#keywords alternative splicing
SUMMARY #length 1896 #checksum 8680

Query Match 98.4%; Score 11009; DB 2; Length 1896;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1547; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db	1	VFSEDLSSLYFNASLQEVFASTGTGLVPCPAAGIPVTLRMVLTAGEIYDVGRH	60
Qy	15	VFSEDLSSLYFNASLQEVFASTGTGLVPCPAAGIPVTLRMVLTAGEIYDVGRH	74
Qy	75	VHPNGTQIRPFPPSSFTLHNDTYCTAENSGKIRSDVIRKAVLEPTVVRVEDK	134
Db	61	VHPNGTQIRPFPPSSFTLHNDTYCTAENSGKIRSDVIRKAVLEPTVVRVEDK	120
Db	121	TMKGNVAFKCIIPSSVEAYITVSMKDTVSLVSGSRFLITSGALYIKDVONDEGLYN	180
Qy	135	TMKGNVAFKCIIPSSVEAYITVSMKDTVSLVSGSRFLITSGALYIKDVONDEGLYN	194
Db	181	YRCITRRHRYGTGRQNSARLFWSDPANSPILDGFDRKAMAGORVELPCALGHPEP	240

```
195 YRCITRRHYTGETFROSNSARLFVSDPANSPASILDGDFHRKAMAGQVLELPCRALCHPEF 254
241 DYRLKXNMPLLEISGRQKTYVTGLLENIRPSDGSVYCVSNRYGAKYIGLYKOPF 300
255 DYRLKXNMPLLEISGRQKTYVTGLLENIRPSDGSVYCVSNRYGAKYIGLYKOPF 314
301 KATISPRKXSVSGOVSLSQSVTGTEDDELXSWYRNGEILNPGKNVARTIGIINHEMLIMDH 360
315 KATISPRKXSVSGOVSLSQSVTGTEDDELXSWYRNGEILNPGKNVARTIGIINHEMLIMDH 374
361 MKVSDGAGYQCFYKDKLSADYVQVYVLEDTGRTIISAFSEKXVSPAEPYSLKCNKGP 420
375 MKVSDGAGYQCFYKDKLSADYVQVYVLEDTGRTIISAFSEKXVSPAEPYSLKCNKGP 434
421 LPTITWTLDDPILKGSGRHSRISQMITSEGNVSVYLNSSQVBDGQVYRGTANNASGVYL 480
435 LPTITWTLDDPILKGSGRHSRISQMITSEGNVSVYLNSSQVBDGQVYRGTANNASGVYL 494
481 YQARINVRGPASIRPMKNITAIAGROTYIHCRAVIGYPYYSIKWYKSNLLPFNHROVAF 540
495 YQARINVRGPASIRPMKNITAIAGROTYIHCRAVIGYPYYSIKWYKSNLLPFNHROVAF 554
541 NNGTLKLSADVOKEDVEGYTCNVLPOLSTSSQSVHTVAVPFIQFEPFEPRESIGORVF 600
555 NNGTLKLSADVOKEDVEGYTCNVLPOLSTSSQSVHTVAVPFIQFEPFEPRESIGORVF 614
601 IPCVVGSGDLPITTTMOCKDGRPIGSLGVTIDNIDFTSSRLSNLSMHNANTCIARNE 660
615 IPCVVGSGDLPITTTMOCKDGRPIGSLGVTIDNIDFTSSRLSNLSMHNANTCIARNE 674
661 AAAVEHOSQLIVRVPKRVVOPRDQDGIYKAVILNCSAGCPYPTIYVWKFSGAGVPOF 720
675 AAAVEHOSQLIVRVPKRVVOPRDQDGIYKAVILNCSAGCPYPTIYVWKFSGAGVPOF 734
721 OPTLNGRIQVLSNGSLILIKHVVEDSGYLYCKVSNVDGAVDSMWLYTKIPAMITSYF 780
735 OPTLNGRIQVLSNGSLILIKHVVEDSGYLYCKVSNVDGAVDSMWLYTKIPAMITSYF 794
781 NNTLATOGOKKEMSCSTAHGEXPIIVRWEKEDRIINPEMARLYSTKEVEGEVISTLOILP 840
795 NNTLATOGOKKEMSCSTAHGEXPIIVRWEKEDRIINPEMARLYSTKEVEGEVISTLOILP 854
841 TVREDSCGFSCSHAINSGEDRGIIQLTVQEPDPPEIEIKDYKARTITLWMTMGFDGNSP 900
855 TVREDSCGFSCSHAINSGEDRGIIQLTVQEPDPPEIEIKDYKARTITLWMTMGFDGNSP 914
901 ITGVDIIECKKNSDSMDSNQRTKDVSPOLNSATIIDHPSSTYSIRMYAKKRICKSEPSNE 960
915 ITGVDIIECKKNSDSMDSNQRTKDVSPOLNSATIIDHPSSTYSIRMYAKKRICKSEPSNE 974
961 LTTTADAEAPDPGPOEYHLEPISSOSIRVTKWAKPKHLQNGIIRGOYIGYREXSTGDNFQ 1020
975 LTTTADAEAPDPGPOEYHLEPISSOSIRVTKWAKPKHLQNGIIRGOYIGYREXSTGDNFQ 1034
1021 FNITSVDTSSGSEYVYLDNLKFTQYGLVQACNRAGTSPSSQOIIITTTLEDVPSYPEN 1080
1035 FNITSVDTSSGSEYVYLDNLKFTQYGLVQACNRAGTSPSSQOIIITTTLEDVPSYPEN 1094
1081 VOALATSPESISISNSTLSKALNGILOGFRVIYWANIMLMOGELIEKNITTTOPSLELDG 1140
1095 VOALATSPESISISNSTLSKALNGILOGFRVIYWANIMLMOGELIEKNITTTOPSLELDG 1154
1141 LEKYTNNSIOVLATFRAGDGVRSQOIFRTKEDVPGPAGVYKAAASASVFWMLPPLK 1200
1155 LEKYTNNSIOVLATFRAGDGVRSQOIFRTKEDVPGPAGVYKAAASASVFWMLPPLK 1214
1201 LNGIIRKTYVCSHPYPIVLSSEFASPDSESYRIPNLSRNRQSVWVVAATSGRGSSSE 1260
1215 LNGIIRKTYVCSHPYPIVLSSEFASPDSESYRIPNLSRNRQSVWVVAATSGRGSSSE 1274
1261 IITVEPLAKAPARILTESGVTTPMKNDIYLPCKAVGDPSPAVVWKMKDSNGTSLVITDG 1320
```

```
1275 IITVEPLAKAPARILTESGVTTPMKNDIYLPCKAVGDPSPAVVWKMKDSNGTSLVITDG 1334
1321 RSISNSGSIIRRTYKABDSGYSCIANNNMGSDIILNLQVYPPQPRLTYSKITSSS 1380
1335 RSISNSGSIIRRTYKABDSGYSCIANNNMGSDIILNLQVYPPQPRLTYSKITSSS 1394
1381 IITLSWLPDNGSGSSIRGYILOYSEDNSEQWGFPIISPEERSYRLNLCGTWYKFTLTAQ 1440
1395 IITLSWLPDNGSGSSIRGYILOYSEDNSEQWGFPIISPEERSYRLNLCGTWYKFTLTAQ 1454
1441 NGVGRIRSEIIEATLTCKEPOFSKEDLFPASINTTRVRLNLIGMNGGCPITSFTLEVR 1500
1455 NGVGRIRSEIIEATLTCKEPOFSKEDLFPASINTTRVRLNLIGMNGGCPITSFTLEVR 1514
1501 PGCTTWTTAQRTSLSKSYILYDLOEATWYELQWRVNCNSAGCAEKQAN 1548
1515 PGCTTWTTAQRTSLSKSYILYDLOEATWYELQWRVNCNSAGCAEKQAN 1562
```

RESULT 2

```
ENTRY 158164 #type complete
TITLE BIG-1 protein - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
```

ACCESSIONS

```
REFERENCE 158164
```

```
#authors Yoshihara, Y.; Kawasaki, M.; Tanii, A.; Tamada, A.; Nagata, S.; Kagamiyama, H.; Mori, K.
```

```
#journal Neuron (1994) 13:415-426
#title BIG-1: a new FGF-1/F3-related member of the immunoglobulin superfamily with neurite outgrowth-promoting activity.
```

```
#cross-references MUID:94338697
```

```
#accession 158164
```

```
##status preliminary; translated from GB/EMBL/DBJ
```

```
##molecule_type mRNA
```

```
##residues 1-1028 #label RES
```

```
##cross-references EMBL:U1091; NID:g563132; PIDN:AAA65607.1; PID:g563133
```

GENETICS

```
#gene BIG-1
```

```
CLASSIFICATION #superfamily contactin; fibronectin type III repeat homology; immunoglobulin homology
```

```
SUMMARY #length 1028 #molecular_weight 112788 #checksum 5866
```

```
Query Match 7.2%; Score 809; DB 2; Length 1028; Best Local Similarity 24.7%; Pred. No. 2,856-135; Matches 250; Conservative 249; Mismatches 447; Indels 66; Gaps 54;
```

```
DB 25 GYVYKPEP-SNIFP--V-GSEYWKIILNCEARGNPSPHYRWOLNGSDIDTSLDRK-L 79
QY 306 GRLYYKQPLKATISPRKXSVSGOVSLSQSVTGTEDDELXSWYRNGELNPGKNVARTIGI 365
DB 80 NCGNLIVINPNRNMOTGSCYOCFA-TNSIGTIVSREAKLOFAYLENFKSRMSRVSVREGQ 138
QY 366 NHENLIMDHMKVS-DGAYQCFVRKDKLSADYVQVYVLEDEG-TKIIISAFSEKY-VSPA 422
DB 139 GYVLLCGPPHSGELSYAVWENEPYFVEEDSRREVSQETGHLIYAKVPSDVGNVTCV 198
QY 423 PSLMKCNVWG-TPLPLTITWTLDDP-LKGSNR-ISO---M-ITS-E-GNVSYLNIS 472
DB 199 TSTVINARVLSPTPLVLRSDCYMGEYEPKIELOFPETLPAKG-STY-K---LECFAL 252
QY 473 SSQVDRGGYV-RCIA-N-NSAGVY-LYQARINVRGPASIRPMKNITAIAGROTYIHCRAV 528
DB 253 GNPVPOINM-RSDDMPEPTK-IKLRFNGVLEIPNFQOE-DTGSYEC-IMENSGRKNVA 308
QY 529 GYPIYSIKMYKKNLPLPNHRGVAFEN-NGTLKLSADVOKEDVEGYTCNVLPQQLSTSQ 587
DB 309 KGRLLTYAKPYWVOLLKDVETAVEDSLYMEGRA-SGRKPSYRWLKNQDAL--VLEERIQ 365
QY 588 SVHVVYKVPF-IQFEPFEPRESIGORVFIQVVGSDLPITITWCKDGRPIGSLGVTID 646
```

 N O T E S

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Apr 25 13:32:01 2000: Maspar time 48.10 Seconds
 975,450 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-956-991-11
 Description: (1-1571) from US08956991A.pep
 Perfect Score: 11189
 Sequence: 1 MRLALSLQSFANVSEDL.....NSAGCAKQAKKARCKFS 1571

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 56.705; Variance 92.824; scale 0.611

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	744	6.6	1447	1 DCC_MOUSE	TUMOR SUPPRESSOR PROTE	1.11e-141
2	740	6.6	1447	1 DCC_HUMAN	TUMOR SUPPRESSOR PROTE	1.06e-140
3	720	6.4	1040	1 AXOI_HUMAN	AXONIN-1 PRECURSOR (AX	8.57e-136
4	703	6.3	1036	1 AXOI_CHICK	AXONIN-1 PRECURSOR (AX	1.24e-131
5	701	6.3	1040	1 AXOI_RAT	AXONIN-1 PRECURSOR (AX	3.84e-131
6	677	6.1	1284	1 NRCA_CHICK	NG-CAM RELATED CELL AD	2.78e-125
7	663	5.9	1010	1 CAMT_CHICK	CONTACTIN PRECURSOR (N	7.14e-122
8	637	5.7	1259	1 CAMT_RAT	NEURAL CELL ADHESION M	1.47e-115
9	637	5.7	1260	1 CAMT_MOUSE	NEURAL CELL ADHESION M	4.19e-114
10	631	5.6	1018	1 CONT_HUMAN	CONTACTIN PRECURSOR (G	4.19e-114
11	631	5.6	1020	1 CONT_MOUSE	CONTACTIN PRECURSOR (N	4.19e-114
12	620	5.5	1257	1 CAMT_MOUSE	NEURAL CELL ADHESION M	1.93e-111
13	609	5.4	1029	1 LAR_MOUSE	PROTEIN-TYROSINE PHOSP	8.76e-109
14	598	5.3	1897	1 LAR_MOUSE	PROTEIN-TYROSINE PHOSP	3.94e-106
15	570	5.1	1837	1 NCN2_MOUSE	NEURAL CELL ADHESION M	2.11e-99
16	574	5.1	1912	1 PTPD_HUMAN	PROTEIN-TYROSINE PHOSP	2.33e-100
17	562	5.0	1837	1 NCN2_HUMAN	NEURAL CELL ADHESION M	2.75e-97
18	558	5.0	1091	1 NCAL_CHICK	NEURAL CELL ADHESION M	1.58e-96
19	562	5.0	1239	1 NRAG_MOUSE	NEUROGLIN PRECURSOR	1.75e-97
20	543	4.9	848	1 NCAL_HUMAN	NEURAL CELL ADHESION M	6.05e-93
21	524	4.7	1088	1 NCAL_XENLA	NEURAL CELL ADHESION M	2.01e-88
22	523	4.7	1092	1 NCAL_XENLA	NEURAL CELL ADHESION M	3.48e-88
23	531	4.7	1266	1 NCAL_CHICK	NEURONAL-GLIAL CELL AD	4.37e-90

24	509	4.5	761	1 NCAL_HUMAN	NEURAL CELL ADHESION M	7.26e-85
25	498	4.5	853	1 NCAL_BOVIN	NEURAL CELL ADHESION M	2.89e-82
26	492	4.4	858	1 NCAL_RAT	NEURAL CELL ADHESION M	7.52e-81
27	496	4.4	1115	1 NCAL_MOUSE	NEURAL CELL ADHESION M	8.57e-82
28	478	4.3	725	1 NCAL_MOUSE	NEURAL CELL ADHESION M	1.48e-77
29	401	3.6	928	1 FAS2_SCHAM	FASCICLIN II PRECURSOR	1.11e-59
30	402	3.6	3707	1 PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	6.53e-60
31	378	3.4	811	1 FS22_DROME	FASCICLIN II, PHOSPHAT	1.94e-54
32	378	3.4	811	1 FS21_DROME	FASCICLIN II, MEMBRANE	1.94e-54
33	364	3.3	4933	1 PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	2.84e-51
34	313	2.8	2477	1 FINE_MOUSE	FIBRONECTIN PRECURSOR	6.17e-40
35	308	2.8	2477	1 FINE_MOUSE	FIBRONECTIN PRECURSOR	7.64e-39
36	313	2.8	2481	1 UN52_CAEL	BASEMENT MEMBRANE PROT	6.17e-40
37	287	2.6	345	1 OPM_MOUSE	OPIOID BINDING PROTEIN	2.69e-34
38	286	2.6	1131	1 MYPE_CHICK	MYOSIN-BINDING PROTEIN	4.41e-34
39	294	2.6	2265	1 FINE_HOVIN	FIBRONECTIN (FN)	8.36e-36
40	284	2.5	345	1 OPM_HUMAN	OPIOID BINDING PROTEIN	1.18e-33
41	279	2.5	345	1 OPM_RAT	OPIOID BINDING PROTEIN	1.39e-32
42	264	2.4	337	1 G55A_CHICK	NEURITE INHIBITOR GP35	2.08e-29
43	267	2.4	1142	1 MYPE_HUMAN	MYOSIN-BINDING PROTEIN	4.88e-30
44	262	2.3	333	1 AMAL_DROME	AMALGAM PROTEIN PRECUR	5.51e-29
45	262	2.3	2386	1 FINE_HUMAN	FIBRONECTIN PRECURSOR	5.51e-29

ALIGNMENTS

RESULT ID	1	STANDARD:	PRT: 1447 AA.
AC	DCC_MOUSE		
AC	P70211:		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.		
GN	DCC.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;		
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BALB/C; TISSUE-BRAIN:		
RX	MEDLINE: 96112625.		
RA	COOPER H.M., ARMES P., BRITTO J., GAD J., WILKS A.F.;		
RT	"Cloning of the mouse homologue of the deleted in colorectal cancer gene (MDC) and its expression in the developing mouse embryo."		
RL	Oncogene 11:2243-2254(1995).		
RN	[2]		
RP	REVISIONS.		
RC	STRAIN-BALB/C; TISSUE-BRAIN:		
RA	COOPER H.M.;		
RL	Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.		
CC	-1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.		
CC	-1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN.		
CC	-1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY ALTERNATIVE SPLICING.		
CC	-1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.		
CC	-1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION AND REMAIN AT THIS LEVEL IN THE ADULT.		
CC	-1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.		
CC	-1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@isb-sib.ch).		

DR EMBL: X85788; CAAS9786.1; -
 DR HSP: P56276; 1TLK-
 DR MGD: MG1.94869; DCC.
 DR PFM: PF00047; 1g; 4.
 DR Glycoprotein: Immunoglobulin domain; Transmembrane; Signal;
 KW Anti-oncogene; Alternative initiation; Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1447
 FT CHAIN 85 1447
 FT CHAIN 26 1097
 FT TRANSMEM 1098 1122
 FT DOMAIN 1123 1447
 FT DOMAIN 54 124
 FT DOMAIN 154 219
 FT DOMAIN 254 317
 FT DOMAIN 345 407
 FT DOMAIN 426 522
 FT DOMAIN 525 618
 FT DOMAIN 619 716
 FT DOMAIN 722 816
 FT DOMAIN 840 940
 FT DOMAIN 941 1042
 FT DOMAIN 61 117
 FT DISULFID 161 212
 FT DISULFID 261 310
 FT DISULFID 352 400
 FT CARBOHYD 60 60
 FT CARBOHYD 94 94
 FT CARBOHYD 298 299
 FT CARBOHYD 318 318
 FT CARBOHYD 478 478
 FT CARBOHYD 628 628
 FT CARBOHYD 702 702
 FT VARSPLIC 819 838
 FT SEQUENCE 1447 AA: 158298 MW: F5FB79BA CRC32:
 Query Match 6.68; Score 744; DB 1; Length 1447;
 Best Local Similarity 29.68; Pred. No. 1,11e-141;
 Matches 228; Conservative 178; Mismatches 293; Indels 72; Gaps 55;

Db 72 IKKKKGLIALGMDRRKQDLPGNSLLIONILHSRHKDEGLYOCASLADSGSIISRT 131
 Oy 535 IKWKKS-NL-LFNNHQAFAENNGTL-K-LSDQKVEDEGYTCNV-LVQPOLSTSS 588
 Db 132 AKYTVAGPLRLFSQTESIAF-MGDTVLKCEYIGEPMP-TIMQNNODL-NPLRGDSR 188
 Oy 589 VHTVVKVP-PFIOPE-FPHFSIGORVFTPCVVSGDLPITIMQDGRPIPSLGVTID 646
 Db 189 VVVLPSGALDISLQPDGSGVYRCSARNPASIRTGNAEVRILSDPGLRQLYLRPSN 248
 Oy 647 NIDETS-SLRISNLNLMHNGNYTCLIANEAA-VEHOSOL-IVRV-PK-FVQ-PRD 698
 Db 249 VIAIEGDAVLECCVSGYPPSEFTW-L-RGEVYQLR-S-K-KYSLGGSNLLISNVD 302
 Oy 699 ODGIYGAAILNCSABEYPTIWKFKSGAGVQFQPIALNGRIOVLSNGSILLIKHVE 758
 Db 303 DSGTYTCVVTYK-NENISASALTYLVPPWFLNHPNLAYMSMDIEFCAVSG-KPV 360
 Oy 759 EDSGYLYCKSNMGVAGVSKMYLTVKIPAMITTSYPTTLATGOKKEMSTHAGEKPT 817
 Db 361 TVNMKNGDVIPS-D-YF-QI-VGG-SNLRLGVKSDGEFYOCVANEAGNAQSS 412
 Oy 818 IVNKEKEIRIINPEMARIYVSTKEVGEVISTDILPTVEDSGFCHAINSVEGRGI 877
 Db 413 AOLIVPRAPSSSILDPAPDVLVLYSSRFVLSWRPAEAKGNIOTTFVFSRGDN 472
 Oy 878 IQLTYQEPDPE-I-EI-KDV-----KARTITLRMTMGDGSPTTGVDIECKNSDS 928
 Db 473 REBALNTQPGS-LQ-LTVGNLKEPAMTFRVAVYNEGPFSSQPIKVAIOPELQVGP 530
 Oy 929 WDSQORTKDVSPOLNSATIIDIHSSSTYSIMYAKNRIGKSEPSNELITLAD-EAARDGP 987

Db 531 VENVLAIVSPTSILITWEPRA-Y-ANGPVQYRLCFTEVSTGKE-Q-NI-EVD--GLS- 582
 Oy 988 POEVLHEPISQSIRVWTKAPKHLQNGIIRGYIGREYSTGCFNFOFIISVTSQDSE 1047
 Db 583 -YKLEIGKETEYTLRFLAIVNRYGPGVSTDDITVTLSDVSPAPONISLEVNSRIKV 641
 Oy 1048 VYTLNLNKFITGYGLVQACNACGSPSSQELITLEDVSPRENVQALATSPESISI 1107
 Db 642 SWLPSPGTCNGFITGYKIRHAKTIRGDM-ETLE-PNNMWL-FTGKSGSOYFOVSA 698
 Oy 1108 SWTLSEKALNGLIOGFRVIYWANLMDGLEIKNTTQPSLEDGLGLEYKTYNSIOVLA 1167
 Db 699 MTYNGTGPSSMYTAEETPNDDESQVDPDPSLHVPRQTCII-MSWTPPLNPIVYRG 757
 Oy 1168 FTRAGGVASE--QIFTRKE-D---VGPAPGAVKAAASAMVSVSLPPLKLGIIIRK 1221
 Db 758 YIIGYGVSPYAEYRV-DSK-QRY-YSTIERLESSHYISLKAFFNNGEG 805
 Oy 1222 YTV-F-CSHPYPTVISEFASPSFSYRIPNLNRNROYSVWVAVTSAGRG 1270

RESULT 2
 ID DCC-HUMAN STANDARD; PRT: 1447 AA.
 AC P43146;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR (COLORECTAL CANCER SUPPRESSOR).
 GN DCC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9501.532.
 RA HEDRICK L., CHO K.R., FEARON E.R., WU T.-C., KINZLER K.W.,
 RA VOGELSTEIN B.;
 RT "The DCC gene product in cellular differentiation and colorectal
 tumorigenesis.";
 RL Genes Dev. 8:1174-1183(1994).
 RN [2]
 RP SEQUENCE OF 1-750 FROM N.A.
 RX MEDLINE: 90100559.
 RA FEARON E.R., CHO K.R., NIGRO J.M., KERN S.E., SIMONS J.W.,
 RA RUPPERT J.M., HAMILTON S.R., PREISINGER A.C., THOMAS G., KINZLER K.W.,
 RA VOGELSTEIN B.;
 RT "Identification of a chromosome 18q gene that is altered in
 colorectal cancers.";
 RL Science 247:49-56(1990).
 RN [3]
 RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBELD EXONS).
 RX MEDLINE: 91121517.
 RA NIGRO J.M., CHO K.R., FEARON E.R., KERN S.E., RUPPERT J.M.,
 RA OLINER J.D., KINZLER K.W., VOGELSTEIN B.;
 RT "Scrambled exons.";
 RL Cell 64:607-613(1991).
 RN [4]
 RP GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
 RX MEDLINE: 94245241.
 RA CHO K.R., OLINER J.D., SIMONS J.W., HEDRICK L., FEARON E.R.,
 RA PREISINGER A.C., HEDGE P., SILVERMAN G.A., VOGELSTEIN B.;
 RT "The DCC gene: structural analysis and mutations in colorectal
 carcinomas.";
 RL Genomics 19:525-531(1994).
 RN [5]
 RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
 RX MEDLINE: 94243823.
 RA MIYAKE S., NAGAI K., YOSHINO K., OTO M., ENDO M., YUASA Y.;
 RT "Point mutation and allelic deletion of tumor suppressor gene DCC in
 human esophageal squamous cell carcinomas and their relation to
 Cancer Res. 54:3007-3010(1994).
 CC -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.

241 RVELPCALGHPEDDYRMLKDNMPELDSGRFOKTVTGLLIENTIRPSPSGSYCEVSNRYG 300
Db 301 TAKYIGLYKOPKATISPRKXSSVGSQVSLSCSVTGEDEDELSYRNGELINGKNV 360
Qy 301 TAKYIGLYKOPKATISPRKXSSVGSQVSLSCSVTGEDEDELSYRNGELINGKNV 360
Db 361 RITGINENLIMDMHVKSDGAYOCFVRKDKLSADYVQVVLDEGTPKIISAESEKVVSP 420
Qy 361 RITGINENLIMDMHVKSDGAYOCFVRKDKLSADYVQVVLDEGTPKIISAESEKVVSP 420
Db 421 AEPVSLMCNVKGTPLPTITWLLDDPILKGGSHRISQMITSEGNVSYLNISSQVARDG 480
Qy 421 AEPVSLMCNVKGTPLPTITWLLDDPILKGGSHRISQMITSEGNVSYLNISSQVARDG 480
Db 481 VYRGTANNSAGVLYOARINVRGPASIRPMKNITAIAGRDYIHCRVIGPYYSIKMYKN 540
Qy 481 VYRGTANNSAGVLYOARINVRGPASIRPMKNITAIAGRDYIHCRVIGPYYSIKMYKN 540
Db 541 SNLLPFNRQVAFENNGITLKSVDQKEVEGEYTCNVLPQSLTSSQSVHYTVKVPFIQ 600
Qy 541 SNLLPFNRQVAFENNGITLKSVDQKEVEGEYTCNVLPQSLTSSQSVHYTVKVPFIQ 600
Db 601 PPEPRFSGRVPFPCVVSQGLPITTTMOKDGRPIGSLGVTIDNIDFTSSLRISNLS 660
Qy 601 PPEPRFSGRVPFPCVVSQGLPITTTMOKDGRPIGSLGVTIDNIDFTSSLRISNLS 660
Db 661 LMHNGNYCIAERNEAAVEHOSQILIVRPKVVOPRQDQIYKAVILNCSAGYEVPT 720
Qy 661 LMHNGNYCIAERNEAAVEHOSQILIVRPKVVOPRQDQIYKAVILNCSAGYEVPT 720
Db 721 IWMKFSKAGVPOFOPIALNGRIQVLSGSLIKHVEEDSGYLLCKVSNVGVADVSKM 780
Qy 721 IWMKFSKAGVPOFOPIALNGRIQVLSGSLIKHVEEDSGYLLCKVSNVGVADVSKM 780
Db 781 YLTIVKIPAMITSYPTTLTATOGOKKEMSCTAHGEKPIIVREKEDRIINPEMAYILVSTK 840
Qy 781 YLTIVKIPAMITSYPTTLTATOGOKKEMSCTAHGEKPIIVREKEDRIINPEMAYILVSTK 840
Db 841 EVGEEVISTLOILPTVRDESGFFSCHAINSXYGDRGIIQLTVQEBPDPPELEIKVKART 900
Qy 841 EVGEEVISTLOILPTVRDESGFFSCHAINSXYGDRGIIQLTVQEBPDPPELEIKVKART 900
Db 901 ITLRMTMGFDGNSPTITGDIIECKNKSQRTKQVSPQNSATITIDHPSYSYSTRM 960
Qy 901 ITLRMTMGFDGNSPTITGDIIECKNKSQRTKQVSPQNSATITIDHPSYSYSTRM 960
Db 961 YAKNRIGKSEPNELITADEAPADGPOEVHLEPISQSIRVTKAKRKHLLONGIIRGY 1020
Qy 961 YAKNRIGKSEPNELITADEAPADGPOEVHLEPISQSIRVTKAKRKHLLONGIIRGY 1020
Db 1021 QIGYREYSTGNGFOENIISVDTSGSEYVTLIDNLKFTQYGLVQACNRAGTGPSSOETI 1080
Qy 1021 QIGYREYSTGNGFOENIISVDTSGSEYVTLIDNLKFTQYGLVQACNRAGTGPSSOETI 1080
Db 1081 TTTLEDEVPYPPENVOAITSPESSISISMTLSKEALNGILOGFVITWAMIMODELCEI 1140
Qy 1081 TTTLEDEVPYPPENVOAITSPESSISISMTLSKEALNGILOGFVITWAMIMODELCEI 1140
Db 1141 KNTTTPSLELDGLEYKTYNSIOVLATFRAGDGVASEQIFFTKEDVPGPAGYKAAA 1200
Qy 1141 KNTTTPSLELDGLEYKTYNSIOVLATFRAGDGVASEQIFFTKEDVPGPAGYKAAA 1200
Db 1201 SASMNVSVSLPPLKNGITIRKTYTVCSPHYPIVISEFEASPDSPSYRIPLNLSRNOYSW 1260
Qy 1201 SASMNVSVSLPPLKNGITIRKTYTVCSPHYPIVISEFEASPDSPSYRIPLNLSRNOYSW 1260
Db 1261 VVAVTISAGGNSSEITVEPLAKAPARILTFSGIVTTPMKMDIVLPCKAVGDPSPAVKMM 1320
Qy 1261 VVAVTISAGGNSSEITVEPLAKAPARILTFSGIVTTPMKMDIVLPCKAVGDPSPAVKMM 1320
Db 1321 KDSNGTSPSLVTTIDGRSISNGSFIIIRYKAEDSGYSCSIANNMGSDEIILNLOVOVPP 1380
Qy 1321 KDSNGTSPSLVTTIDGRSISNGSFIIIRYKAEDSGYSCSIANNMGSDEIILNLOVOVPP 1380

Qy 1321 KDSNGTSPSLVTTIDGRSISNGSFIIIRYKAEDSGYSCSIANNMGSDEIILNLOVOVPP 1380
Db 1381 DQPLRVSKTSSITLWMLPDNGSSIRGYILQYSEDNSEQSGSPISPSERSYLEN 1440
Qy 1381 DQPLRVSKTSSITLWMLPDNGSSIRGYILQYSEDNSEQSGSPISPSERSYLEN 1440
Db 1441 LKCGTWYKFTTLAONGVPGRISEIIEKTLGKRPQSKDELFASINTRYRLNLGN 1500
Qy 1441 LKCGTWYKFTTLAONGVPGRISEIIEKTLGKRPQSKDELFASINTRYRLNLGN 1500
Db 1501 DGGCPTSFTELEPRFGTTWTMTAORTSLSKSYLLIDQEAATWYELOMRVNCSAGCAEQ 1560
Qy 1501 DGGCPTSFTELEPRFGTTWTMTAORTSLSKSYLLIDQEAATWYELOMRVNCSAGCAEQ 1560
Db 1561 AKEARCKEFS 1571
Qy 1561 AKEARCKEFS 1571
RESULT 2
ID 060468 PRELIMINARY: PRT: 1896 AA.
AC 060468;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE (FRAGMENT).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA YAMAKAWA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
RA LYONS G.E., KORENBERG J.R.;
RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF023449; AAC17966.1; -
DR PFAM: PF00041; fn3: 6.
DR PFAM: PF00047; lg: 8.
FT NON_TER 1
SO SEQUENCE 1896 AA; 209785 MW; 0DEGEQCE CRC32:
Query Match 98.4%; Score 11009; DA 4; Length 1896;
Best local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1547; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 1 VSEEDLHSSLIYFNASLOEVYFVASTTGTIVLPCPAGIPPVTLRWTLATGELIYDVPGIRH 60
Qy 15 VSEEDLHSSLIYFNASLOEVYFVASTTGTIVLPCPAGIPPVTLRWTLATGELIYDVPGIRH 74
Db 61 VHPNGTLQIFPPPPSFSFTLHUNTYCTAENPSGKIRSQDVHKAIVLREPTYVVEDOK 120
Qy 75 VHPNGTLQIFPPPPSFSFTLHUNTYCTAENPSGKIRSQDVHKAIVLREPTYVVEDOK 134
Db 121 TMRGNVAFKCIIPSSVEAYITVVSWEKDTVLSVSGRFLITSGALYIKDYQNEDEGLYN 180
Qy 135 TMRGNVAFKCIIPSSVEAYITVVSWEKDTVLSVSGRFLITSGALYIKDYQNEDEGLYN 194
Db 181 YRCTRHRRTGEMROSARLFVSDPANSAPSTLDGFDHRKMAAGRVLELPCALGHPER 240
Qy 195 YRCTRHRRTGEMROSARLFVSDPANSAPSTLDGFDHRKMAAGRVLELPCALGHPER 254
Db 241 DYRWLKDMPLELSGRFOKTVTGLLIENTIRPSPSGSYCEVSNRYGTAKVIGRLYKQPL 300
Qy 255 DYRWLKDMPLELSGRFOKTVTGLLIENTIRPSPSGSYCEVSNRYGTAKVIGRLYKQPL 314
Db 301 KATISPRKXSSVGSQVSLSCSVTGEDEDELSYRNGELINRPGKNRITGINENLIMDH 360
Qy 315 KATISPRKXSSVGSQVSLSCSVTGEDEDELSYRNGELINRPGKNRITGINENLIMDH 374
Db 361 MVKSDGAYOCFVRKDKLSADYVQVVLDEGTPKIISAESEKVVSPAEBVSYLMCNVKGTP 420
Qy 375 MVKSDGAYOCFVRKDKLSADYVQVVLDEGTPKIISAESEKVVSPAEBVSYLMCNVKGTP 434

AVTSAGRNSSEITVEPLAKAPARILFESGVTTPMKKDIVLPCKAVGDSPPAVKW
KSNKGTPLSVITIDGRSRISFISNGSFIIPTVKAEDSGYVCIANNMGSDSEIILNLOVY
PPDQRLRYTSKTTSSITISLRLPDNGSSLRGYLQIYSEDNSEMGSPFISPEBSY
RLBNLKCWTWFTLTAQNGVGRHISIEIKTKGKPKPQSKDEULFASINTVRVL
NLIGNNDGCCPIITSEFLRPPGTIVMTIATQTSLSKSYILYDLOEATWYELQWVNC
SAGCAEKQKAEARCKEFS"

BASE COUNT 1633 a 1779 c 1709 g 1292 t
ORIGIN

alignment_scores:

Quality: 8223.00 Length: 1571
Ratio: 5.234 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-08-956-991-11 x AF023450 ..

Align seg 1/1 to: AF023450 from: 1 to: 6413

1 MetTPIleLeuAlaLeuSerLeuPheGlnSerPheAlaAsnValPheSe 17
453 ATGTGATACAGGCTCTCTCTGTCAGAGCTTCGGAATGTTTTCAG 502
17 TGIAspLeuHisSerSerLeuTyrPheValAsnAlaSerLeuGlnLuv 34
503 TGAAGACCTACACTCCAGCCCTTCTTGTCAATGCATCTCTGCAAGG 552
34 a1Val1PheAlaSerThrThrGlyThrLeuValProCysProAlaAlaGly 50
553 TAGTGITTCACACACACGCGGACTGTGTCCTGCCCGCAGCAGGC 602
51 IleProProValThrLeuArgTyrPheValaThrGlyGlnGlyLeu 67
603 ATCCCTCTGTGACTGTACATGATGATACCTACCGGCGAGGATCTA 652
67 rAspValProGlyIleArgHisValHisProAsnGlyThrLeuGlnIleP 84
653 CGATGTCCCGGATCCGCGACGTCCACCCCAAGCGCATCTCCAAATT 702
84 heProPheProProSerSerPheSerThrLeuIleHisAspAsnThrTyr 100
703 TCCCTTCCCTCTTCAAGCTTCAAGCTTAATCATGATTAATACTAT 752
101 TyrCysThrAlaGluAsnProSerGlyLysIleLeuSerGlnAspValH 117
753 TATTGCACAGCTGAHAATCCTTCAGGAAATTAAGAGTCAGAGATGTCCA 802
117 s1LeLysAlaValLeuArgGluProTyrThrValArgValGluAspGln 134
803 CATCAAGGCTGTTTACGGAGCCCTAATACAGTCCGTGTGAGAGACAGA 852
134 ystHMetArgGlyAsnValAlaValPheLysCysIleIleProSerSer 150
853 AAACCATGAGAGCAATGTCCGGTCTTCAAGTGCATTATCCCTCTCG 902
151 ValGluAlaTyrIleThrValIleSerTyrPgluLysAspThrValSerLe 167
903 GTGAGAGCGTACATCACTGCTGCTCATGGAGAAACACACTGTTTACT 952
167 uValSerGlySerArgPheLeuIleThrSerThrGlyAlaLeu TyrIleL 184
953 TGCTCTCAGATCTAGATTCTTCATCATCATCCAGCGAGCCTGTATATTA 1002
184 ysAspValGlnAsnGluAspGlyLeuTyrAsnTyrArgCysIleThrArg 200
1003 AAGATGTACAAATGAAGATGATGTATTAATACCGTGCATCACGCGG 1052
201 HisArgTyrThrGlyLuvThrArgGlnSerAsnSerAlaArgLeuPheVa 217
1053 CATGATATACACGAGAGACGAGCAGCAGCAGCAGCAGCAGCAGCTTTTGT 1102
217 lSerAspProAlaAsnSerAlaProSerIleLeuAspGlyPheAspHis 234

1103 ATCAACACCCAGGAACTACAGCCCATCATCTACTGATGGTTTACCATC 1152
234 rGlyAlaMetAlaGlyIleAlaArgValGluLeuProCysLysAlaLeuGly 250
1153 GCAAGACCATGGCTGGCAGCGCTGTGAGAGTGCCTTGCAAGCCGTCGGG 1202
251 HisProGluProAspTyrArgTyrLeuLysAspAsnMetProLeuGluLe 267
1203 CACCTGAGCCAGATTACCGCTGCTGAAGAGCAACATGCCCCGGAAC 1252
267 uSerGlyArgPheGlnLysThrValThrGlyLeuLeuIleGluAsnIleA 284
1253 TTCAGGAGAGTCCAGAAAGACCGTACGCGGGCTGCTCATTAAGAACAATC 1302
284 rGProSerAspSerGlySerTyrValCysGluValSerAsnArgTyrGly 300
1303 GCCCTCGGACTCAGCAGCATATGTTGTGAAGTGTCCACACAGATACGGA 1352
301 ThrAlaLysValIleGlyArgLeuTyrValLysGlnProLeuLysAlaTh 317
1353 ACTGCTAAGGTGATAGCCCGCTGACTGTAACAGCCACTGAAAGCCAC 1402
317 rIleSerProAlrGlyValLysSerSerValGlySerGlnValSerLeus 334
1403 CATCAGTCCAGAGAGGTAAAGACGCGTGGTAGCCCAATCTTCCTTGT 1452
334 erCysSerValThrGlyThrGluAspGlnGluLeuSerTyrTrArgAsn 350
1453 CCTGAGCGGTGACAGAACTGAGCCAGCAACTCTCTGTGTACCGCAAT 1502
351 GlyGluIleLeuAsnProGlyLysAsnValArgIleThrGlyIleAsnH 367
1503 GGTGAATTCCTCAACCTCGAATAAATGTGAGATCACAGGATCAACCA 1552
367 sGluAsnLeuIleMetAspHisMetValLysSerAspGlyGlyAlaTyrG 384
1553 CGAAACCTTATATGATGATCATGTGTCAAAGTGACGGGGCCCATACC 1602
384 lncCysPheValArgLysAspLysLeuSerAlaGlnAspTyrValGlnVal 400
1603 AGTGCTTTGTGGCAAGAGCAAGCTGTCCGCTCAAGACTAATGTCCAGGTG 1652
401 ValLeuGluAspGlyThrProLysIleIleSerAlaPheSerGluLysVa 417
1653 GTCTTGGAAGATGAACTCCCAAAATTATTTCTGCTTTAGTGAAAGGT 1702
417 lValSerProAlaGluProValIleSerLeuMetCysAsnValLysGlyThrP 434
1703 GGTAGTCCAGCAGACCGCGGTTTCCCTTATGTCCACAGCTGAAGGAACAC 1752
434 roLeuProThrIleThrTrpThrLeuAspAspAspProIleLeuLysGly 450
1753 CTTTGCCACGATTCACGTGAGCCCTGAGCATGACCCGATTTCTCAAGGT 1802
451 GlySerHisArgIleSerGlnMetIleThrSerGluGluAsnValIse 467
1803 GGCATCACCGCATCAGCCAGATGATCAGTCCGAGGGGGAACGTGGTCAG 1852
467 rTyrLeuAsnIleSerSerSerGlnValArgAspGlyGlyValTyrArgC 484
1853 CTACCTGAACATCTCCAGCTCCAGAGTCCGAGCGGGGAGTATACCGCT 1902
484 ystThrAlaAsnAsnSerAlaGlyValValLeuTyrGlnAlaArgIleAsn 500
1903 GCACGTGCCAAACACTCGCGGGAGTGTCTGTACCAAGGCTCGAATAAC 1952
501 ValArgGlyProAlaSerIleArgProMetLysAsnIleThrAlaIleAl 517
1953 GTAAGAGGCGCTGCAAGCATTCGACCAATGAAAAACATCACAGCAATATGC 2002
517 aGlyArgAspThrTyrIleHisCysArgValIleGlyTyrProTyrTyrS 534

17 rGluAspLeuHisSerSerLeuTyrPheValAsnAlaSerLeuGlnGluV 34
|||||
503 TGAAGACCTTACTCCAGCCTTACTTGTTCATGTCATCTCTCCAGAGG 552
34 aValPheAlaSerThrThrGlyThrLeuValProCysProAlaIleGly 50
|||||
553 TAGGTTCGCAGCACCAGGGGACTGTGGTCCCTCCCGCAGCGAGGC 602
51 IleProProValThrLeuArgTrpTyrLeuAlaThrGlyGlnGluIleTy 67
|||||
603 ATCCCTCTGTGACTCTCAGTGTGATGCTTACCCAGCGGCGAGAGATCTA 652
67 AspValProGlyIleArgHisValHisProAsnGlyThrLeuGlnIleP 84
|||||
653 CGATGTCCCGGGATCCGCCAGCTCCACCCCAAGCGCATCTCCAAATTT 702
84 heProPheProPheSerSerPheSerThrLeuIleHisAspAsnThrTyr 100
|||||
703 TCCTCTCCCTCTTCAAGCTTCAGTACCTTAATCCATGATTAATCTTAT 752
101 TyrCysThrAlaGluAsnProSerGlyLysIleArgSerGlnAspValHi 117
|||||
753 TATTTCACAGCTGAAATCTTCAGGGAAATTAGAAGTCAGATGTCCA 802
117 sIleLysAlaValLeuArgGluProTyrThrValArgValGlnAspGln 134
|||||
803 CATCAAGCTGTATTACGGAGCCCTATACAGTCCGTGTGGAGAGCCAGA 852
134 ystHmeArgGlyAsnValAlaValPheLysCysIleIleProSerSer 150
|||||
853 AAACCATGAGAGCAATGTGGCTCTTCAAGTCATTAATCCCTCCCTCG 902
151 ValGluAlaTyrIleThrValValSerTrpGluLysAspThrValSerLe 167
|||||
903 GTGGAGGCGTACATCAGTCTCTCATGAGGAGAAAGACACTGTTCCT 952
167 uValSerGlySerArgPheLeuIleThrSerThrGlyAlaLeuYrIleI 184
|||||
953 TGTCTAGGATCTAGATTCTCATCCATCCACGGAGCCCTGTATATA 1002
184 yAspValGlnAsnGluAspGlyLeuTyrAsnTyrArgCysIleThrArg 200
|||||
1003 AACATGTACAGATGAAGATGGATTGTATTACTACCGCTGCATCCAGCG 1052
201 HisArgTyrThrGlyLutThrArgGlnSerAsnSerAlaArgLeuPheVa 217
|||||
1053 CATCGATACACCGAGAGAGAGGAGGACGACGACGAGCCAGACTTTTGT 1102
217 lSerAspProAlaAsnSerAlaProSerIleLeuAspGlyPheAspHisA 234
|||||
1103 ATGAGACCCAGCACTCAAGCCCATCCATACGTGATGGTTGACCATTC 1152
234 rGlyAlaMetAlaGlyGlnArgValGluLeuProCysLysAlaIleGly 250
|||||
1153 GCAAAAGCATGGCTGGGCAAGCGGTGTGAGAGTGCCTTGCAAAAGCGCTGGG 1202
251 HisProGluProAspTyrArgTrpLeuLysAspAsnMetProLeuGlnLe 267
|||||
1203 CACCTTGAGGCAAGATTACCGCTGGCTGAAGACACATGCCCCCTGGAAGT 1252
267 uSerGlyArgPheGlnLysThrValThrGlyLeuLeuIleGlnAsnIleA 284
|||||
1253 TTCAGGAGGTTCCAGAAAGCGTGAAGGGGCTGCTCATTTGAGAAACATTC 1302
284 rqrProSerAspSerGlySerTyrValCysGluValSerAsnArgTyrGly 300
|||||
1303 GCCCTTCGGACTCAGGCACTATGTTGTGAAGTGTCCACACATACGGA 1352
301 ThrAlaLysValIleGlyArgLeuTyrValLysGlnProLeuLysAlaTh 317
|||||
1353 ACTGTAAGGTGATAGCGCCGCTGTACGTGAACACGCACTGAAGGCAC 1402

317 rIleSerProAlaLysValLysSerSerValGlySerGlnValSerLeus 334
|||||
1403 CATCAGTCCAGAGAAAGTTTAAACACAGCTGGTAGCCAACTTCTCTGT 1452
334 erCysSerValThrGlyThrGluAspGlnGluLeuSerTrpTyrArgAsn 350
|||||
1453 CCTGACGGGTGACAGAACTGAGGACCGACGAACTCTCTGCTACCGCAAT 1502
351 GlyGluIleLeuAsnProGlyLysAsnValArgIleThrGlyIleAsnHi 367
|||||
1503 GGTGAATCCTCAACCTCGAAGAAAAATGTGAGATCAACAGGATCAACCA 1552
367 sGluAsnLeuIleMetAspHisMetValLysSerAspGlyValAlaTyrG 384
1553 CGAAACCTTATATGATGCATGCATGTCAAAAGTGACGGGGCGCATACC 1602
384 lncSspHeValArgLysAspLysLeuSerAlaGlnAspTyrValGlnVal 400
1603 AGTGCTTGTGCGCAAGCAAGCTGTCCGCTCAAGACTATGTGCGAGTG 1652
401 ValLeuGluAspGlyThrProLysIleIleSerAlaPheSerGluLysVa 417
|||||
1653 GTCTTGAAAGATGAAGACTCCAAATTAATTTCTGCTTAGGAAAGAGT 1702
417 lValSerProAlaGluProValSerLeuMetCysAsnValLysGlyThrP 434
1703 GGTGAGTCCAGCAGACCGCTTCCCTTATGTGCAACGTGAAGGGAACAC 1752
434 rLeuLeuProThrIleThrTrpThrLeuAspAspProIleLeuLysGly 450
1753 CTTTGCCACAGATCAGTGAACCTGAGCGATGACCCGATTTTCAAGGT 1802
451 GlySerHisArgIleSerGlnMetIleThrSerGluLysAsnValValSe 467
1803 GGCAGTCACCGCATTCAGCCAGATGATCAGTCGAGGGGAAGCTGCTCAG 1852
467 rTyrLeuAsnIleSerSerSerGlnValArgAspGlyValValTyrArgC 484
1853 CTACCTGAAACATCTCAGCTCCAGGTCCGGAGCGGGGAGTCTACCGCT 1902
484 ystHraIaAsnAsnSerAlaGlyValValLeuTyrGlnAlaArgIleAsn 500
1903 GCACGTGCACAACTCGCGCGAGTGTCTCTGTACCAAGCTCGAATTAAC 1952
501 ValArgGlyProAlaSerIleArgProMetLysAsnIleThrAlaIleAl 517
1953 GTAGAGGCGCTGCAAGCATTCGACCAATGAATAACATCACACCAATAC 2002
517 acGlyArgAspThrTyrIleHisValArgValIleGlyTyrProTyrTyrS 534
2003 AGGAGGGACACATCACTTCACTGTCTGTGATGGCTATACCGTATTACT 2052
534 erIleLysTrpTyrLysAsnSerAsnLeuLeuProPheAsnHisArgGln 550
2053 CCATTAAATGGTACAAAGACTTAACCTGCTCTTCAACCAACCCCA 2102
551 ValAlaPheGluAsnAsnGlyThrLeuLysLeuSerAspValGlnLysG 567
2103 GTGCAATTTGAACAATGGAACCTTAAACTTTCAGATGTGCAAAAGGA 2152
567 uValAspGluGlyLutTyrThrCysAsnValLeuValGlnProGlnLeus 584
2153 AGTGACAGAGGGGAGTACACAGTGCACAGTGTGTTCAACCAACACTCT 2202
584 erThrSerGlnSerValHisValThrValLysValProProPheIleGln 600
2203 CCACAGCCAGAGCGTCCACGTGACGCTGAAGAGTTCCGCTTTCATACAA 2252
601 ProPheGluPheProArgPheSerIleGlyGlnArgValPheIleProCy 617
2253 CCCTTGAGTTTCAAGATCTCCATTGGGACGCGGTCTTCACTCCCTG 2302
617 sValValValSerGlyAspLeuProIleThrIleThrTrpGlnLysAspG 634

149 CCTGACAGATGTTGGACACGACCTCTACTTTGTAATGACTCTTGACG 198
33 luvAlvalPheAlaSerThrThrGlyThrLeuValProCysProAlaAla 49
AGGTACCTTTTCCAGCTCCGTGGGGGTGTGTGCTCCGCCGCGCG 248
50 GLYILEProValThrLeuArgTrpTyrLeuAlaThrGlyGluIle 66
GGCTGCCCGCCACGGCCCTTCGATGTAAGTGGCCACAGGGGACGACAT 298
249 GGCTGCCCGCCACGGCCCTTCGATGTAAGTGGCCACAGGGGACGACAT 298
66 eTYrAspValProGlyIleArgHisValHisProAsnGlyThrLeuGln 83
CTACAGCTGCCCGCACATCCGGACGTCACGCCACAGCGACGCTGACG 348
299 CTACAGCTGCCCGCACATCCGGACGTCACGCCACAGCGACGCTGACG 348
83 lPheProPheProPheSerSerPheSerThrLeuIleHisAspAsnThr 99
TCTACCCCTCTCCCTCCGCTTCAATACCTTATCCACGACAATGAC 398
349 TCTACCCCTCTCCCTCCGCTTCAATACCTTATCCACGACAATGAC 398
100 TYrTYrCysThrAlaGluAsnProSerGlyLysIleArgSerGlnAspVal 116
TACTTCTGCACCGCGAGACGCTGCCGCAAGATCCGAGC TCCACAT 448
399 TACTTCTGCACCGCGAGACGCTGCCGCAAGATCCGAGC TCCACAT 448
116 lHisIleLysAlaValLeuArgGluProTYrThrValArgValGluAspG 133
CCGCGTCAAGACAGTTTTCAGGGAACCTACACCGCTCCGGTGAGGATC 498
449 CCGCGTCAAGACAGTTTTCAGGGAACCTACACCGCTCCGGTGAGGATC 498
133 lNlyThrMetArgLysAlaValAlaValPheLysCysIleIleProSer 149
AAAGTTCATCGTGGCAAGCTGGCGCTTCAAGTGCCTCATCCCTCT 548
499 AAAGTTCATCGTGGCAAGCTGGCGCTTCAAGTGCCTCATCCCTCT 548
150 SerValGluAlaTYrIleThrValValSerTrpGluLysAspThrVal 166
TCAGTGCAGGAATATGTTAGCGTTGTATCTGGGGAAGAACACAGTCTC 598
549 TCAGTGCAGGAATATGTTAGCGTTGTATCTGGGGAAGAACACAGTCTC 598
166 lLeuValSerGlySerArgPheLeuIleThrSerThrGlyAlaLeuTYrI 183
CATCATCCACAGAAAACAGTTTATTATCTACCTACACCGCGCTGTACA 648
599 CATCATCCACAGAAAACAGTTTATTATCTACCTACACCGCGCTGTACA 648
183 lLysAspValGluAsnGluAspGlyLeuTYrAsnTYrArgCysIleThr 199
TCTCTGACGTACAGAGAGAGAGCGCTCTCCACCTATCGCTGATCAC 698
649 TCTCTGACGTACAGAGAGAGAGCGCTCTCCACCTATCGCTGATCAC 698
200 ArgHisArgTYrThrGlyGluThrArgGlnSerAsnSerAlaArgLeuP 216
AAGCAGCATATAGCGGGAGACCCGCGAGCATAGGGGACGCCCTCTC 748
699 AAGCAGCATATAGCGGGAGACCCGCGAGCATAGGGGACGCCCTCTC 748
216 eValSerAspProAlaAsnSerAlaProSerIleLeuAspGlyPheAsp 233
TGTGACAGACCTCTGAGTCCGATCCACCATCTCGATGGTTCCACT 798
749 TGTGACAGACCTCTGAGTCCGATCCACCATCTCGATGGTTCCACT 798
233 lSArgLysAlaMetAlaGlyGlnArgValGluLeuProCysLysAlaLeu 249
CCCAGAGAGTGTGGCGCGCCACACCGGAGCTGCCCTGACCGCGCTCG 848
799 CCCAGAGAGTGTGGCGCGCCACACCGGAGCTGCCCTGACCGCGCTCG 848
250 GlyHisProGluProAspTYrArgTrpLeuLysAspAsnMetProLeuG 266
GCCTACCTATCCCGCCATCCGCTGCTCAAGATGGCCGCGCCCTCC 898
849 GCCTACCTATCCCGCCATCCGCTGCTCAAGATGGCCGCGCCCTCC 898
266 uLeuSerGlyArgPheGlnLysThrValThrGlyLeuLeuIleGluAsnI 283
GGCTGCAGCGCTGACCAAGCCATCACAGGCTGACCATCAGCAGCT 948
899 GGCTGCAGCGCTGACCAAGCCATCACAGGCTGACCATCAGCAGCT 948
283 lLeuArgProSerAspSerGlySerTYrValLysGluValSerAspArgTYr 299
TGGGACCGAGAGACGCGCACCTACATTGTGAGTCAACCAACACCTTC 998
949 TGGGACCGAGAGACGCGCACCTACATTGTGAGTCAACCAACACCTTC 998
300 GlyThrAlaLysValIleGlyArgLeuTYrValLysGlnProLeuLysAl 316
GGTTCGAGAGAGCCACAGCATCTCATGTCTATGATCCCTTCATGT 1048
999 GGTTCGAGAGAGCCACAGCATCTCATGTCTATGATCCCTTCATGT 1048
316 aThrIleSerProArgLysValLysSerSerValGlySerGlnValSerL 333
GACCCTGACACCAAGAGCTGAGAGACCGGCATTGGCAGCACGGTCTATCC 1049
1049 GACCCTGACACCAAGAGCTGAGAGACCGGCATTGGCAGCACGGTCTATCC 1098

333 euSerCysSerValThrGlyThrGluAspGlnGluLeuSerTrpTYrArg 349
TCTCTGTCCCTCTGACGGGCTCCCGAGATTCCACCATCGCTGTGATGC 1148
1099 TCTCTGTCCCTCTGACGGGCTCCCGAGATTCCACCATCGCTGTGATGC 1148
350 AsnGlyGluIleLeuAsnProGlyLysAsnValArgIleThrGlyLeuAs 366
AACACGAGACTGTGCTGCTGTACAGAGGTCATCTCATCCCTGAG 1198
1149 AACACGAGACTGTGCTGCTGTACAGAGGTCATCTCATCCCTGAG 1198
366 nHisGluAsnLeuIleMetAspHisMetValLysSerAspGlyAlaArg 383
CAACGAGAGCTGTCTCATCTGCTGCCAGAGAGCCATTCCGGGGCT 1248
1199 CAACGAGAGCTGTCTCATCTGCTGCCAGAGAGCCATTCCGGGGCT 1248
383 yGlnCysPheValArgLysAspLysLeuSerAlaGlnAspTYrValGln 399
ACCAAGCTTCTGCTACCCGCAAGGCCACAGCCCGCCAGACTTTCCTC 1298
1249 ACCAAGCTTCTGCTACCCGCAAGGCCACAGCCCGCCAGACTTTCCTC 1298
400 ValValLeuGluAspGlyThrProLysIleIleSerAlaPheSerGlyL 416
ATTGCACCTGAGAGATGGCAGCGCCGACATCGTCTGCTTCCAGGAA 1348
1299 ATTGCACCTGAGAGATGGCAGCGCCGACATCGTCTGCTTCCAGGAA 1348
416 sValValSerProAlaGluProValSerLeuMetCysAsnValLysGlyT 433
GGTGTCAACCCCGGGGAGCAGTCTCATCTGATGTGCGGCAAGGCG 1398
1349 GGTGTCAACCCCGGGGAGCAGTCTCATCTGATGTGCGGCAAGGCG 1398
433 hPheLeuProThrIleThrTrpThrLeuAspAspProIleLeuLys 449
CCCCGCCGCCCGGCTACCTGAGCTGGCCCTCGACGATGACCCCATCTG 1448
1399 CCCCCGCCGCCCGGCTACCTGAGCTGGCCCTCGACGATGACCCCATCTG 1448
450 GlyGlySerHisArgIleSerGlnMetIleThrSerGlyGly 463
GATGGAGCGCCACCGCACCAACAGTACACCATGTGCGACGCG 1490
1449 GATGGAGCGCCACCGCACCAACAGTACACCATGTGCGACGCG 1490

seq_name: /cgn2_6/p/tdata/1/ina/PCTUS9_COMB.seq:PCT-US94-05277-1

seq_documentation_block:
Sequence 1, Application PC/TUS9405277
GENERAL INFORMATION:
APPLICANT: Bruskin, Arthur
APPLICANT: Jarosz, David E.
APPLICANT: Johnson, Kate
APPLICANT: Kinzler, Kenneth W.
APPLICANT: Vogelstein, Bert
APPLICANT: Zaretsky, James R.
TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05277
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,42709
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202,508,9100
TELEFAX: 202,508,9299
TELEX: 197430 BBWB UT
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4608 base pairs

OM of: US-08-956-991-11 to: EST: * out_format : pfs
 Date: Apr 25, 2000 9:52 PM

About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framer-p2n.model -DEV=xlp
 -O=/cgn2.1/USPTO.spool/US08956991/runat.25042000.12.5.7.9550/app-query.fasta.1
 -DB=EST -QPM=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
 -MIMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELETE=7.000 -START=1 -MATRIX=blonum62 -TRANS=human40.cdi
 -LIST=45 -LOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFM=pts -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08956991
 -NCPU=6 -ICPU=3 -NO_XLPEXT -WAIT -THREADS=1

Search information block:

Query: US-08-956-991-11
 Query length: 1571
 Database: EST: *
 Database sequences: 4538634
 Database length: 1887831982
 Search time (sec): 2119.170000

Sequence	Strid	Orig	Zscore	Escore	Len	Documentation
gb_est27:AI454704	-	723.00	1426.94	2.3e-72	458	AI454704 UI-R-BT0-qk-h-03-0-UI.
gb_est18:AI5159035	+	665.00	1338.23	9.0e-66	427	AM159035 zas0e05.x1 Xenopus EST
gb_est18:AI003472	-	604.00	1215.47	6.2e-59	355	AJ003472 AJ003472 Selected chrc
gb_gssi:FR0002941	+	576.00	1150.03	2.7e-55	619	Z86724 F. rubripes GSS sequenc
gb_est18:AI106884	+	411.00	810.29	2.3e-36	646	AI106884 GH06134.5prime GH Dros
gb_est18:AM048129	-	408.50	812.87	1.9e-36	402	AM048129 UI-R-BH1-alcq-t-04-0-UI
gb_est18:AA697364	+	408.50	801.37	6.3e-36	786	AA697364 HL02328.5prime HL Dros
gb_est18:AI404957	+	355.50	658.02	4.1e-30	569	AI404957 GH26371.5prime GH Dros
gb_gssi:FR0021997	+	346.50	678.51	5.0e-29	610	AI404957 F. rubripes GSS sequenc
gb_est18:AI542894	+	346.50	674.61	8.2e-29	696	AI542894 SD09407.5prime SD Dros
gb_est17:AM014717	+	344.00	675.99	4.6e-29	409	AM014717 UI-R-BT0-aae-a-07-0-UI
gb_est12:AA325603	+	333.50	659.93	5.4e-28	298	AA325603 EST88915 Cerebellum II
gb_gssi:FR0030733	+	329.50	648.17	2.4e-27	442	AL027102 Fugu rubripes GSS sequ
gb_gssi:FR0002934	+	328.00	645.25	3.6e-27	437	Z86717 F. rubripes GSS sequenc
gb_est18:AI541596	+	317.50	618.36	1.1e-25	635	AI541596 SD02777.5prime SD Dros
gb_gssi:FR0021994	-	298.50	579.99	1.5e-23	605	AI014865 F. rubripes GSS sequenc
gb_gssi:FR0030753	+	293.50	569.51	5.9e-23	614	AL027122 Fugu rubripes GSS sequ
gb_est18:AI406159	+	289.50	565.70	9.6e-23	450	AI406159 tm82h11.x1 NCI_CGAP_BI
gb_gssi:FR0030758	-	285.00	552.72	5.1e-22	585	AI406159 GH26371.5prime GH Dros
gb_gssi:FR0002938	-	275.50	532.79	6.5e-21	602	AL027127 Fugu rubripes GSS sequ
gb_gssi:CN8010PH	-	274.00	532.94	1.6e-20	785	AL099263 Drosophila melanogast
gb_est13:AI107879	-	272.00	525.96	1.6e-20	587	AI107879 GH05755.5prime GH Dros
gb_est18:AI519340	+	250.00	480.96	5.0e-18	578	AI519340 LD38885.5prime LD Dros
gb_gssi:FR0022007	+	246.50	477.54	7.8e-18	443	AL014678 F. rubripes GSS sequenc
gb_est18:AI543452	+	246.00	472.94	1.4e-17	570	AI543452 SD10207.5prime SD Dros
gb_est18:AV264594	+	240.00	470.59	1.9e-17	282	AV264594 AV264594 RIKEN full.1e
gb_gssi:CN800058	+	231.50	433.79	2.1e-15	1101	AL050788 Drosophila melanogast
gb_est10:AA812746	+	226.50	427.76	4.6e-15	816	AA812746 LD22445.5prime LD Dros
gb_gssi:FR0022012	+	224.50	428.80	4.1e-15	549	AL014883 F. rubripes GSS sequenc
gb_gssi:CN801040	+	224.00	419.78	1.3e-14	937	AL098490 Drosophila melanogast
gb_est18:AI5131274	+	222.50	422.79	8.8e-15	649	AI5131274 SD02242.5prime SD Dros
gb_gssi:FR0009351	-	221.00	440.79	8.7e-16	147	X88325 H. sapiens DNA for trapp
gb_est14:AM213058	+	221.00	433.95	2.1e-15	238	AL000636 F. rubripes GSS sequenc
gb_gssi:CN801293	+	220.00	420.48	1.2e-14	532	AM213058 un98g10.y1 NCI_CGAP_M
gb_est18:AI513189	+	215.00	401.81	1.3e-13	1033	AI103165 Drosophila melanogast
gb_est18:AI513189	+	215.00	408.19	5.7e-14	613	AI513189 LD5391.5prime LD Dros
gb_est18:AI513189	+	215.00	408.05	7.5e-14	711	AI513189 LD5391.5prime LD Dros
gb_gssi:FR0002080	+	214.50	406.85	7.2e-14	544	AL004083 AU004083 Bombyx mori F
gb_est12:AI722009	+	214.00	406.45	7.0e-14	595	AI722009 fd18c09.y1 Sugano Kawa
gb_gssi:CN80070A	+	208.00	386.46	9.2e-13	1028	AI067489 Drosophila melanogast
gb_gssi:AO815550	+	207.00	395.52	2.9e-13	470	AO815550 HS_5346_A2_B08_SPEE RH

gb_gssi:FR0030755 + 206.50 393.93 3.5e-13 489 ! AL027124 Fugu rubripes GSS s
 gb_est18:AI515655 + 206.50 393.64 3.7e-13 499 ! AI515659 LD2472.5prime LD D
 gb_est18:AL119288 + 203.50 384.88 1.1e-12 599 ! AL119288 DXF2P61A081_r1 761

seq_name: gb_est127:AI454704

seq_documentation_block:

LOCUS AI454704 458 bp mRNA EST 05-JUL-1999
 DEFINITION UI-R-BT0-qk-h-03-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone
 UI-R-BT0-qk-h-03-0-UI 3', mRNA sequence.

ACCESSION AI454704

VERSION AI454704.1 GI:4295587

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Scurionathia; Muridae; Murinae; Rattus.

1 (bases 1 to 458)

Normalization and subtraction: two approaches to facilitate gene
 discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

MEDLINE

97044477 On Mar 20, 1998 this sequence version replaced g1:2980496.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Oligo-dr track not found. Not a site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.
 Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
 through Research Genetics (www.resgen.com) This clone is also
 available through the I.M.A.G.E. Consortium at LNL
 (info@image.llnl.gov). IMAGE ID=1788947

Seq primer: M13 Forward

FEATURES

1..458 Location/Qualifiers
 CDS join=1
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BT0-qk-h-03-0-UI"
 /clone_lib="UI-R-BT0"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; This library
 (UI-R-BT0) consists of a mixture of individually tagged
 normalized libraries constructed from rat hippocampus,
 thalamus, mid-brain, medulla, corpus striatum, cerebral
 cortex and testis. The tag used to identify the source
 tissue is a string of 3-6 nucleotides present
 Not I site and the oligo-dr track which allows
 identification of the library of origin of a clone within
 the mixture. This library was then subtracted using a
 driver consisting of a mixture of all clones from UI-R-A0,
 UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and
 UI-R-C2p."

BASE COUNT

107 a 112 c 123 g 116 t

ORIGIN

alignment_scores: 723.00 Length: 152
 Ratio: 5.021 Gaps: 0
 Percent Similarity: 94.737 Percent Identity: 91.447

alignment_block:

US-08-956-991-11 x AI454704/rev ..

FEATURES	Location/Qualifiers
source	1..427
	/organism="Xenopus laevis"
	/db_xref="taxon:8355"
	/clone="za50e05"
	/clone_lib="Xenopus EST library"
	/tissue_type="total brain tissue"
	/cell_line="M2-TGA"
	/dex_stage="tadpole"
	/note="Vector: Lambda Zap I; Site: 1: XbaI; This library was supplied by Holly Cline (Cold Spring Harbor Labs). cDNA synthesis with oligo df Xba I (Xba I cloning site) RNA: stage 50-56 tadpoles, total brain tissue, GTC extraction method."
BASE COUNT	114 a 122 c 91 g 100 t
ORIGIN	
alignment_scores:	
Quality:	665.00 Length: 139
Ratio:	4.890 Gaps: 1
Percent Similarity:	97.842 Percent Identity: 92.806
alignment_block:	
US-08-956-991-11 x AM159035	..
Align seg 1/1 to: AM159035 from: 1 to: 427	
410	lleseralpheserglulysvalyserproalaglprovalserle 426
17	ATCTCGCATTTAGCGAGAACTGTGTGACCCCGGAGAACCTGTGCTCCCT 66
426	umetCysasnVallysGlyThrProleuProthrllethrTrpThrleua 443
67	TATGTGCACAGTAAAGAGACCCCTCTTCCGACATCCTTGGACGCTGG 116
443	spaspsProilleuLysGlySerHisArgIleserGlnmetlle 459
117	ACGACGACCCCATTTATCAAGAGCGCCACACAGATAAGTCAGTATATC 166
460	ThrserGlnGlyAsnValyserIleuasnIleserSerSerGlnva 476
167	ACCTCTGAGGCGAACGTTGTCACTTACCTGGAACATCACCACTCAAGT 216
476	largaSpGlyGlyValTyArgCysThrAlaAsnAsnserAlaGlyVal 493
217	CGAGATGGCGGGGCTACCGATGCACCGCTAACACACTCTCGGGCGCTG 266
493	alleuTyGlnAlaArgIleAsnValArgGlyProAlaSerIleArgPro 509
267	TCTTTTACCAAGCTCGAATAAAGCT.....CTGCAGATATCCGTCCA 309
510	metIysAsnIleThrAlaIleAlaGlyArgAspThrTyrlleHisCysAr 526
310	ATGAAAAACATCACGGCCATAGCTGTGGGGGAGACATATCATCTGCGG 359
526	gValIleGlyTyProTyTySerIleLysTrpTyLysAsnserAsnL 543
360	CGTCATTGGATTCCTCTACTACTCATCATGATGAGACAAAACCTTAAC 409
543	euleuProPheAsnHis 548
410	TCCTCCCATTTAATCAC 426
seq_name:	gb_est18:AJ003472
seq_documentation_block:	
LOCUS	AJ003472 355 bp mRNA EST 01-APR-1998
DEFINITION	AJ003472 selected chromosome 21 cDNA library Homo sapiens cDNA
ACCESSION	AJ003472
VERSION	AJ003472.1 GI:2769503
KEYWORDS	EST.